JULY / AUGUST 2001

THE JOURNAL OF HIGHER EDUCATION

13

Models of Institutional Resource Allocation: Mission, Market, and Gender
Cindy S. Volk, Sheila Slaughter, and Scott L. Thomas

The Ties That Bind: Presidential Involvement with the Development of NCAA Division I Initial Eligibility Legislation
Dan Covell and Carol A. Barr

No Rookies on Rookies: Compliance and Opportunism in Policy Implementation
Michael R. Mills and Adrienne E. Hyle

Quality Assessment of University Students: Student Perceptions of Quality Criteria
Janet G. Donald and D. Brian Denison

COR-H05850-13-HP1107

JE

The Journal of Higher Education

LEONARD L. BAIRD, *Editor*

Editorial Board

JAMES FAIRWEATHER, *Chairperson,*
Michigan State University

ESTELA BENSIMON, *University of Southern California*

SYLVIA HURTADO, *University of Michigan*

GARY R. PIKE, *University of Missouri-Columbia*

SHEILA A. SLAUGHTER, *University of Arizona*

FRANCES STAGE, *New York University*

CAROLINE S. TURNER, *University of Minnesota*

AMI ZUSMAN, *University of California*

DONNA BROWDER EVANS, *Dean, College of Education*
The Ohio State University, ex officio

EDWARD J. RAY, *Executive Vice President*
and Provost, The Ohio State University, ex officio

YOLANDA MOSES, *President, American Association*
for Higher Education, ex officio

LEONARD L. BAIRD, *The Ohio State University, ex officio*

Copyright © 2001 by The Ohio State University
All Rights Reserved

The Journal of Higher Education (ISSN 0022-1546) Published bimonthly by Ohio State University Press. Ava M. Stinnett, Managing Editor; Bertina Povenmire, Production Editor; Publishing and advertising offices: Ohio State University Press, 1070 Carmack Road, Columbus, Ohio 43210-1002. Periodicals postage paid at Columbus, Ohio, and at additional mailing offices. POSTMASTER: Send address changes to: *The Journal of Higher Education,* Ohio State University Press, 1070 Carmack Road, Columbus, Ohio 43210-1002.

The Paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials. ANSI Z39.48-1992.

Paper Subscription. Libraries and other institutions, $90.00; individuals, $42.00; association members, $28.00; students, $28.00. Outside the U.S. add $7.00 postage. **Electronic Subscriptions** are available to participating libraries and institutions through OCLC's FirstSearch. Orders and remittances should be sent to Journal Subscription Manager, Ohio State University Press, 1070 Carmack Road, Columbus, OH 43210-1002. Claims on undelivered copies must be made within six months of the date of publication. Allow six weeks for address change.

Back Issues, Reprints, Microfilms. Please contact the publisher for availability of back issues. Reprints of individual articles, reprints of out-of-print issues, and microfilms of back volumes are available from University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106, 734-761-4700. On the Web at www.umi.com. Back issues from 1930 are available online to participating libraries and institutions via JSTOR. On the Web at www.jstor.org.

Manuscripts. Articles submitted for publication should be mailed in triplicate to Leonard L. Baird, Editor, *The Journal of Higher Education,* Ohio State University Press, 1070 Carmack Road, Columbus, OH 43210-1002. Manuscripts should meet the criteria outlined in the Instructions to Contributors. On the Web at www.ohiostatepress.org/journals/jhecontr.htm.

Abstracts and Indexes. *Abstracted in Academic Abstracts, Academic Librarianship, Educational Administration Abstracts, ERIC Clearinghouse on Higher Education, Higher Education Abstracts, Information Access Co., Political Science Abstracts, Psychological Abstracts, Readers Guide to Periodical Literature, Social Science Source, Social Work Research and Abstracts,* and *Society for Research into Higher Education. Indexed in Arts & Humanities Citation Index, Bibliographic Index, Book Review Index, Education Index,* and *Social Sciences Citation Index.*

Copyright © 2001 by The Ohio State University. All rights reserved. Article copying exclusively for personal or internal use, or for the personal use of specific clients, is permitted provided that the copy user or producer pay a fee of $2.00 per copy of any article, comment, book review, or other complete unit to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. This limited permission does not extend to other kinds of copying, such as copying for general distribution, for resale, for creating new collective works, or for advertising and promotion purposes. Specific written permission for such copying must be obtained from Ohio State University Press. Code: 0022-1546/2001 $2.00 + 0.

Cover: D. Boone/Corbis

16/8/2001

JHE July/August 2001
Vol. 72, No. 4

The Journal of Higher Education

CINDY S. VOLK, SHEILA SLAUGHTER, and SCOTT L. THOMAS 387

Models of Institutional Resource Allocation: Mission, Market, and Gender

Two major models of resource allocation were tested to determine which best explained allocation of state dollars. Using multiple regression, we showed that departments characterized by (male), full-time faculty, graduate degrees, and grants and contracts tend to be more highly resourced than those characterized by (female) faculty, high (female) adjunct use, and undergraduate student credit hours.

DAN COVELL and CAROL A. BARR 414

The Ties That Bind: Presidential Involvement with the Development of NCAA Division I Initial Eligibility Legislation

This article provides a chronology of presidential efforts to deal with the conflicts related to the reconciliation of academic mission and athletic success through the development of NCAA initial eligibility academic legislation and analyzes these efforts in terms of maintaining congruence within the constituency-based American higher education system.

MICHAEL R. MILLS and ADRIENNE E. HYLE 453

No Rookies on Rookies: Compliance and Opportunism in Policy Implementation

This case study analyzes implementation of an institutional policy restricting the use of graduate teaching assistants in freshmen-level instruction. Findings suggest units viewed implementation as a hierarchically determined compliance task or were able to make implementation an opportunity for creative problem definition and solving akin to the organized anarchy model.

JANET G. DONALD and D. BRIAN DENISON 478

Quality Assessment of University Students: Student Perceptions of Quality Criteria

University students, asked to judge criteria for their own quality, rated a commitment to learning most important of twenty-five criteria. They considered most criteria of student quality to be more important at graduation than at entry, supporting a talent development approach to assessment rather than a resource approach.

BOOK REVIEWS

University Teaching: International Perspectives, edited by James J. F. Forest
REVIEWED BY LESLIE ADAMS WIMSATT 503

Condemning Students to Debt: College Loans and Public Policy, edited by Richard Fossey and Mark Bateman
REVIEWED BY SCOTT R. SWEETLAND 506

Copyright © 2001 by The Ohio State University

JHE September/October 2001
Vol. 72, No. 5

The Journal of Higher Education

PATRICK T. TERENZINI, ALBERTO F. CABRERA,
CAROL L. COLBECK, STEFANI A. BJORKLUND,
and JOHN M. PARENTE 509

**Racial and Ethnic Diversity in the Classroom:
Does It Promote Student Learning?**

Defendants of affirmative action are asked to demonstrate that race-sensitive admissions policies and diversity produce educational benefits for all students. This study indicates that the racial/ethnic composition of a classroom may indeed be related to the development of students' problem-solving and group skills. The nature of that effect, however, may be curvilinear.

DOUGLAS T. SHAPIRO 532

**Modeling Supply and Demand for Arts and Sciences Faculty:
What Ten Years of Data Tell Us About the Labor Market
Projections of Bowen and Sosa**

The labor market projections of Bowen and Sosa's *Prospects for Faculty* (1989) are assessed by testing their assumptions about faculty supply and demand against data from the last decade. Improvements to the model are recommended, including the use of an inventory-attrition model to account for backlogs of supply.

LAMONT FLOWERS, STEVEN J. OSTERLIND, ERNEST T.
PASCARELLA, and CHRISTOPHER T. PIERSON 565

**How Much Do Students Learn in College?
Cross-Sectional Estimates Using the College BASE**

Analyzing data from fifty-six four-year colleges, this investigation employed a cross-sectional design with statistical controls for confounding influences to estimate the net impact of college on five standardized, objective measures of student learning: English, mathematics, science, social studies, and a composite of all four tests. Women in the sample demonstrated significantly smaller net freshman-senior differences across all five measures.

LAURA W. PERNA 584

The Relationship Between Family Responsibilities and Employment Status Among College and University Faculty

This study uses data from the 1993 National Study of Postsecondary Faculty to examine the extent to which the concentration of women among part-time and nontenure-track faculty is related to family responsibilities. Descriptive and multinomial logistic regression analyses are used to examine the research questions.

REVIEW ESSAY

LARRY KEIG 612

The Role of Faculty in Assessing Colleagues' Teaching
Review of *The Peer Review of Teaching: A Sourcebook,*
by Nancy Van Note Chism

BOOK REVIEWS

Enhancing Student Learning: Setting the Campus Context,
edited by Frances Stage, Lemuel W. Watson, and Melvin Terrell
REVIEWED BY PETER M. MAGOLDA 620

Prioritizing Academic Programs and Services: Reallocating Resources to Achieve Strategic Balance, by Robert C. Dickeson
REVIEWED BY DON G. CREAMER 622

Shattering the Myths: Women in Academe, by Judith Glazer-Raymo
REVIEWED BY FLORENCE A. HAMRICK 625

Copyright © 2001 by The Ohio State University

JE November/December 2001
Vol. 72, No. 6

The Journal of Higher Education

MARVIN W. PETERSON and MARNE K. EINARSON 629

**What Are Colleges Doing About Student Assessment?
Does It Make a Difference?**

Using national survey data, this study examines how postsecondary institutions have approached undergraduate student assessment, the organizational and administrative patterns used to support assessment, and the institutional uses and impacts realized from assessment efforts. Patterns of assessment approaches, organizational and administrative support, and uses and impacts differ by institutional type.

KENNETH A. FELDMAN, CORINNA A. ETHINGTON, and JOHN C. SMART 670

**A Further Investigation of Major Field and Person-Environment Fit:
Sociological Versus Psychological Interpretations of Holland's Theory**

Students whose dominant personality type was incongruent with the academic environment in which they majored gained about equally in abilities and interests pertinent to that environment as did students whose dominant personality type was congruent. Results are interpreted in terms of a socialization dynamic being more prevalent than a personality dynamic. Comparisons are made with an earlier, related study that presented ostensibly contradictory results.

NIRA HATIVA, RACHEL BARAK, and ETTY SIMHI 699

**Exemplary University Teachers: Knowledge and Beliefs
Regarding Effective Teaching Dimensions and Strategies**

Four exemplary teachers were interviewed, videotaped in class, and rated by students on a list of effective classroom behaviors. Findings suggest that each teacher achieved excellence using a different set of effective teaching dimensions and strategies but also that they all share a small number of teaching dimensions and strategies.

JOHN ROBST 730

Cost Efficiency in Public Higher Education Institutions

This study reexamines the revenue and cost structures of higher education institutions. First, the article documents the reduced importance of state appropriations and the increased importance of tuition revenues during the early 1990s in public universities. Second it considers how the changed revenue structure influenced cost efficiency in public higher education institutions.

BOOK REVIEWS

Why Universities Matter: A Conversation About Values, Means, and Directions, edited by Tony Coady
REVIEWED BY MARC CUTRIGHT 751

ANNUAL INDEX 754

Copyright © 2001 by The Ohio State University

Cindy S. Volk
Sheila Slaughter
Scott L. Thomas

Models of Institutional Resource Allocation

Mission, Market, and Gender

In an era when expansion is no longer the obvious solution for dealing with curricular change in research universities, the question of how and why institutions allocate resources among departments, the organizational units that deliver curricula, should become increasingly important. Resource constraint has caused widespread restructuring in public research universities, but few studies examine its effects on departments, although the broad goals of restructuring are to redesign institutions to lower costs, achieve greater student learning, give more attention to teaching, and contribute to regional economic development (Guskin, 1994; Gumport & Pusser, 1996; Massey & Zemsky, 1994), all efforts that depend on departmental cooperation. Institutional resource allocation affects departments in a number of ways. The resources available to departments shape who is hired, how much and whom they teach. Quality of faculty and work load, in turn, influence research norms and productivity. Changes in patterns of resource allocation among departments are critical to understanding the shape of knowledge in the twenty-first century.

In the 1980s, prompted by periodic state and institutional budget crises, researchers began to study internal resource allocation among departments within colleges and universities (Ashar, 1987; Ashar & Shapiro, 1990; Hackman, 1985; Melchiori, 1982; Morgan, 1983). But in the 1990s, questions about internal resource allocation among depart-

Cindy S. Volk is adjunct assistant professor, Department of Special Education, Rehabilitation, and School Psychology, The University of Arizona; Sheila Slaughter is professor of higher education, The University of Arizona; Scott L. Thomas is associate professor of higher education administration, University of Hawaii at Manoa.

The Journal of Higher Education, Vol. 72, No. 4 (July/August 2001)
Copyright © 2001 by The Ohio State University

ments generally were put in the broader context of restructuring. This shift meant that attention turned away from departments as units that created and delivered particular kinds of knowledge and curricula efficiently and effectively and toward individual faculty performance on productivity measures as well as individual faculty response to institutional incentives (Layzell, 1996; Levin, 1991). If departments were considered at all, they were treated as generic departments that reacted to institutional and professional incentives rather than as departments organized around concrete kinds of knowledge, peopled by faculty with similar characteristics who trained students for specific careers (Fairweather, 1996; Massy & Wilger, 1992; Massy & Zemsky, 1994). Costing studies were the exceptions, but these used econometric models to identify abstract cost structures among departments aggregated in broad fields of study, a process that reified rather than explained patterns of difference in institutional investment (Brinkman, 1990; Dundar & Lewis, 1994).

Along with a relatively small number of researchers using critical theory and feminist perspectives, we see this shift from department to individual, or, conversely, broad field of study (science and engineering, humanities, social sciences) as the unit of analysis, as masking increasing stratification between departments and within universities (Bellas, 1994, 1997; Gumport, 1993; Kerlin & Dunlap, 1993; McElrath, 1992; Slaughter, 1993; Volk, 1995). Critical researchers view faculty and the curricula they deliver as organized in departments that powerfully mediate individual faculty performance. Faculty delivering some curricula may not receive the same resources as faculty associated with other curricula, just as faculty in departments preparing students for certain careers may not be given the same support as faculty in other, more favored departments. Department, rather than individual faculty performance, may be a powerful explanatory variable. Faculty located in underresourced, overextended departments may not be able to respond fully to the complex array of institutional incentives and disincentives that characterize multimission public research universities. We think that understanding which departments receive resources and why is critical to understanding the far-reaching reorganization and revaluing of knowledge that is presently occurring.

To explore our questions about resource allocation among departments, we reviewed the several theories that purport to explain variance in allocation. We identified two major perspectives, the rational/political and the critical/political. We decided to see which of the theories best explained resource allocation at a single public Research I university (Carnegie Foundation for the Advancement of Teaching, 1994), examining all departments except those located in the medical and law schools.[1] We identified a total of 30 independent variables that represented key components of the explanations offered by the two theories. We looked at standard variables researchers have traditionally used to explain variation in resource allocation among departments such as salary, size, grant dollars, and student FTE. Because we looked at critical/political as well as rational/political factors, we incorporated variables not usually used in departmental resource allocation studies, for example, variables for ascribed characteristics such as gender and minority status of faculty and students. We also developed variables that were proxies for closeness of departments to external markets, for example, assistant professor faculty entry salaries. Finally, we examined variables indicative of internal stratification such as lower division, upper division, and graduate student credit hours. We used multiple regression models to assess the effects of our independent variables on the dependent variable, internal allocation of state dollars to departments. After presenting our results, we discuss what our findings from the final model mean for the two theories and for the restructuring of universities, the reorganization of knowledge, and the stratification of academic work within public research universities.

Theory

The theories that guide researchers who study internal institutional resource allocation among departments generally fall into one of two frames: rational/political or critical/political. Rational/political theorists highlight the role of a small number of decision makers at the apex of the institution and emphasize the functional use of resources to maintain and enhance institutional efficiency and effectiveness (Morgan, 1983). Although they emphasize rationality, rational/political theorists are sensitive to political dimensions of internal budgetary processes and allow for interest group bargaining, acknowledge power differentials among departments, especially in the ability to win external grants and contracts, and recognize that groups external to the university can sometimes intervene to influence allocation (Ashar, 1987; Ashar & Shapiro, 1990; Hackman, 1985; Morgan, 1983). However, rational/political theorists generally explain resource allocation among departments by productivity and meritocratic criteria: funds flow to departments that are central to mission and workload, are productive in terms of student credit hours, grants and contracts and faculty scholarship, and are high in quality. Rational/political theorists acknowledge politics as an important factor, but never see politics as determining resource allocation to the point where they overwhelm institutional economy and efficiency.

The rational/political model initially drew heavily on organizational theory that treated institutions as self-contained and was concerned primarily with internal priority setting. As higher education continued to restructure (but not downsize, see Leslie & Rhoades, 1995) and economics became the preferred social science idiom, the market was incorporated into rational/political models in the form of human capital theory (Breneman, 1988; Slaughter & Leslie, 1997; Youn, 1989). This purportedly linked external labor markets for individual professional skills more closely to internal academic reward systems. Because human capital theory assumed markets to be free, rational and impartial, they were invoked by rational/political theorists to explain and justify increasing differences in salaries by department (Lee, Leslie, & Olswang, 1987).

In contrast, critical/political[2] theorists see resource allocation to departments as shaped by gender, race, power and service to external constituencies dominant in the broader political economy (Bellas, 1994,1997; Gumport, 1993; Kerlin & Dunlap, 1993; McElrath, 1992; Slaughter, 1993; 1998; Volk, 1995). Like rational/political theorists, critical/political theorists see politics and power as playing a part, but critical/political theorists foreground politics and power, arguing that differentials in political economic power among departments explain the majority of variance in resource allocation among departments. They read political power as patterned by gender, race, and relation of departments to federal and corporate research markets, high-end private sector markets for professionals, and the social welfare function of the state (Slaughter, 1997, 1998). Critical/political theorists make the case that concentrating institutional resources on departments close to federal and corporate research markets as well as high-end private sector markets for professionals loads institutional resources on departments where there are many graduate students and few undergraduates, creating systemic productivity problems for public research universities (Volk, 1995). They argue that marked inequities in resources among departments place faculty in some departments—particularly those with heavy lower-division teaching loads—at a disadvantage in responding to the complex array of institutional incentives and disincentives characteristic of multimission public research universities. Because the departments with the least resources and the highest undergraduate teaching loads are often departments with a relatively high number of women and minority faculty, critical/political theorists point out that this pattern of resource allocation undermines some institutional missions and goals, for example, the undergraduate teaching mission and the goal of equity.

Data and Methods

Our study included all departments (72) within a public research university, excluding the medical school and law school. The amount of state appropriated dollars to each department was the dependent measure.[3] We developed various measures on each of the 70 remaining departments. Data defining these variables were gathered from departmental profiles developed and maintained by the Office of Institutional Research, from the Sponsored Projects Office, from the Office of Student Affairs, from the Affirmative Action/Equal Opportunity Office, and from a 1992 University-Wide Quality Review, commissioned by the provost, of the centrality and quality of all departments.

Rational/Political Variables

Like Ashar (1987), Ashar & Shapiro (1990), and Hackman (1985), we did not focus on full professor salaries, a measure that traditionally explains most of the differences among departments with regard to state dollar allocations. We wanted to avoid having our analysis dominated by "pipeline" issues that suggest women and minorities are in the system, moving up, and in time and without intervention will receive salaries similar to men. To address this we treated the salaries of assistant professors as a control, allowing us to concentrate on variables that explained differences among departments, net of the primary cost driver, academic salary. Although we acknowledge the predictive value of full professor salaries, including the salaries of professors at different academic ranks in our model is statistically problematic, forcing us to choose one level over another for inclusion. We chose to include salaries at the assistant as opposed to full level to better capture departments' closeness to market. The rationale for this choice is discussed in a subsequent section. This control allowed us to then focus on the four areas commonly used by rational/political theorists to measure the rational decision-making processes in resource allocation: centrality, workflow, grant and contract productivity, and quality (Ashar 1987; Ashar & Shapiro, 1990; Hackman, 1985).

Centrality. In the literature, centrality usually refers to the importance of departments to university mission or the criticality of departments to maintaining institutional workflow. Centrality to mission (Hackman, 1985) foregrounds departments that teach large numbers of students yet also captures departments that teach few students but are critical to our substantive understanding of the educational task of universities. For example, English departments are central because they teach large numbers of undergraduates, but so are philosophy and classic departments, because faculty commonly agree that a university would not be a university without philosophy and classics departments even though these departments do not have large numbers of students. Centrality (CENTRAL) was measured by using the centrality score assigned to departments by the 1992 University-Wide Quality Review Committee. Faculty committees explicitly charged with judging centrality to mission reviewed extensive documentation submitted by the various departments and assigned them scores of 0 (fails to meet criteria), 1 (meets criteria), and 2 (exceeds criteria).

Workflow. A second rational/political variable was departmental workflow productivity or workflow centrality (Ashar, 1987; Ashar & Shapiro, 1990). Workflow productivity sees departments as central even if they graduate few students, so long as they have a service function through which they offer courses to large numbers of students. For example, mathematics departments often do not graduate large numbers of majors but offer service courses such as introductory mathematics, calculus, and trigonometry. We measured departmental workflow productivity six ways: total student credit hours (TOTALSCH), the proportion of student credit hours taught at the lower division (PCTLDSCH), the proportion of student credit hours taught at the upper division (PCTUDSCH), the proportion of student credit hours taught at the graduate level (PCTGRSCH), number of full-time equivalent (STUFTE) students, and number of full time equivalent faculty (FACFTE).

Grants and contracts. A third rational/political variable was grant and contract productivity (Ashar, 1987; Hackman, 1985) as measured by numbers of external dollars generated by the various departments (GRANTS). Grants and contracts comprised roughly one-third of the budgets of research universities and accounted for 34% of the annual operating budget of the institution under study. Researchers working from a rational/political frame saw departments with strong grants and contracts records as a rational choice for resource allocation because such investment strengthens departments' abilities to keep institutional operating budgets afloat. Institutions value departments' abilities to secure grants and contracts, especially in an era of federal decline in Research and Development (R&D) and volatility within federal R&D programs. However, critical theorists point out that not all departments have the same rich array of external resources available to them; those in science, math, engineering, and the health sciences are favored by federal grant and contract programs (Slaughter, 1997).

Department quality. Departmental quality, a fourth rational/political variable, utilized two proxy measures. We again used external dollars in grants (GRANTS), which were generally highly competitive and peer reviewed, as one measure. The other measure was the quality score (QUALITY) assigned to each department by the 1992 University-Wide

Quality Review Committee. As in the case of the centrality variable, faculty committees reviewed departmental materials and then rated departments with scores of 0 (does not meet criteria), 1 (meets criteria), and 2 (exceeds criteria). Like the grants and contract variable, the quality score is also a research productivity score. Number of publications and awards was the primary indicator in determination of quality rankings. The quality score is our only publication indicator because the university did not keep data on departmental or individual publication records.

Critical/Political Variables

Diversity. The primary critical/political variable was diversity of department. Diversity was measured four ways: numbers of male and female students (MALESTU and FEMSTU) by major, which we converted to departments; faculty gender (FEMFAC), using only full time equivalent faculty; ethnicity ("white" or other) of students by major (MINFAC), converted to departments; ethnicity ("white" or other) of faculty. For each measure, we calculated percentages (PCTFFAC and PCTMFAC) as well as raw numbers. Critical/political theorists argue that ascribed characteristics such as race or gender are reflective of economic and power divisions in the wider society and that they also play a part in determining who gets what in academe. They would expect departments with high numbers of women and minority faculty to receive fewer resources than others. Contrarily, rational/political theorists would make the case that resource allocation patterns are gender neutral because economy and efficiency are value-free and/or more important than ascribed characteristics.

Faculty and student resources. Secondary critical/political variables were the ratio of faculty to FTE faculty lines (FTERATIO) and an FTE student to FTE faculty line ratio (STUFACFT). The FTERATIO captures the use of off-track faculty, adjuncts, and departments' use of faculty from other departments in relation to full-time faculty within the department. Critical/political theorists would expect "have-not" or low-resourced departments to make use of large numbers of adjuncts or off-track faculty as cheap labor to produce SCH and argue that high use of adjunct faculty penalized departments by creating a heavier work load for-full time faculty who have correspondingly increased supervisory roles and committee assignments, reducing their ability to engage in other, more highly rewarded activities, such as research and grant and contract writing. Similarly, critical/political theories would expect high student to faculty ratios (STUFACFT) to result in decreased institutional resource allocations, because faculty in departments with high student to faculty ratios would not have time to engage in graduate education, research, and grant and contract activity.

Other Independent Variables

Closeness to market. Because rational/political theorists are beginning to incorporate elements of human capital and labor market theory into their explanations of resource allocation to departments, we wanted to include variables that measured departments' closeness to the market. Human capital theorists take the position that the greater the market demand for particular types of expertise embodied in graduates, the higher their market value (Thurow, 1995). Therefore, the higher the salary for graduates, the greater the market demands for them, the higher the salary for the faculty who teach them, and the greater the resource allocation to the departments who teach them. Unpacked, the logic of this sequence of argument is as follows. Greater market demand for particular types of expertise drives up salaries in these fields. Graduates of such fields—for example, graduates in advanced computing—are offered many jobs at high starting salaries. To compete for candidates with these skills, universities must offer high salaries, especially for entry level positions. Given that salaries constitute the greatest share of departmental resources, departments that have to compete for faculty in fields with high external market demand need more resources.

Academic labor market theories offer[4] a somewhat different approach to market demand for expertise. Because they see demand for expertise as mediated by organizations, labor market theorists try to explain how organizational preferences order and shape supply and demand in specific markets (Spaeth, 1979; Stolzenberg, 1975; Tolbert, Patrick, & Beck, 1980). Their focus on market segmentation offers a challenge to human capital and rational/political theories. However, research on labor markets for academics looks neither at departments nor the concrete persons who work in departments, but instead focuses on allocation of faculty among various sectors (research universities, comprehensive universities, four-year colleges, community colleges), type of control (private versus public), and strata (elite versus non-elite) (Breneman & Youn, 1988; Tolbert, 1986; Youn, 1989). Nor does academic labor market theory examine empirically competition between universities and corporations for faculty who possess high-demand expertise, the crux of the rational/political and human capital arguments for differential salaries and, by extension, differential institutional resource allocation to departments.

Although feminist theorists concerned with academic labor deal with the same problems as labor market theorists, their work on faculty rarely intersects. Feminists, who are committed to understanding the salience of ascribed characteristics such as gender and ethnicity, include these variables in their approach to academic labor markets (Bellas, 1994, 1997; Tolbert, 1986), while academic labor market theorists generally do not (Youn, 1989).[5] Generally, feminist theorists found, all else equal, that "faculty in fields where women concentrate are paid less than comparable faculty in fields with comparable labor market conditions but with few women employees, . . . indicating discipline level discrimination" (Bellas, 1994, p. 816; see also Bellas, 1997). However, the Bellas studies, by far the most comprehensive to date, did not look at labor market segmentation, specifically at markets for research university faculty, nor did they look at institutional allocation to departments, which included more than salaries, nor at what we call "closeness to the market," which we take to be departments' position relative to the dynamic, high technology, private sector of the market, position relative to which usually predicts the highest salaries (Brint, 1994).

Although we were unable to look at competition among universities for research faculty because we studied only a single institution, we did develop variables for closeness to the market. Because we could locate no direct measures of closeness to the market, we used three proxy measures. The first was the average salary of baccalaureate recipients in their first jobs by department, which we refer to as student exit salary. We reasoned that high salaries would indicate high market demand. We could not find institutional data on exit salary, so we turned to the College Placement Council's national, annual salary summary of initial salary offers by curriculum area. We converted curriculum area to departments and used this information to calculate the average exit salaries of undergraduates by department (GRADSAL). The second proxy measure was initial salary of assistant professors (SALASST). In most studies of faculty compensation, seniority and stints in administration are the variables that most strongly predict high professorial salaries (Konrad & Pfeffer, 1990). These variables are internal to the institution and do not reflect free market competition. Assistant professor salaries are most likely to capture the value of professors on the open market because new PhDs have the option of taking a faculty or private sector position. We collected assistant professor salary data and computed average salaries for assistant professors by department. As a third measure we again used the number of grant and contract dollars generated by the various departments (GRANTS). We thought successful competition for grants and contracts measured the proximity of departments to federal research markets, which, in turn, might signal market competitiveness of departments in relation to research intensive corporations.

Many of the variables—grants, closeness to the market, student credit hours—are used to measure multiple constructs and figure in both rational/political and critical/political theories. For example, rational/political theorists argue that grants are measures of economic viability, scholarly productivity, and quality. They are an economic measure in that they bring external resources to the department, a productivity measure in that they indicate scholarly output, and a quality measure in that they are competitive and refereed. Critical/political theorists do not necessarily disagree with the rational/political theorists' line of argument, but would offer a different interpretation. They view grants as measures of economic productivity and quality but also see them as indicators of departmental closeness to federal research markets, proximity to which provides departments greater power in negotiating for internal resource allocations. Critical/political theorists also make the case that grants are only one of an array of quality measures and perhaps not the one that should be most heavily weighted, particularly if institutions are concerned with teaching and the production of undergraduate student credit hours or degrees. Because these variables are polyvalent, they acquire meaning only in the context of other variables deployed in complex arguments, which we hope to make in our subsequent discussion.

The data were gathered in Academic Year 1992–93. At that time, state allocations to the research university under study mirrored broad national trends, which were characterized by cut-backs at the state level (El Khawas, 1992; Gumport, 1993). We specified a multiple regression model to test the relationships implicit in the aforementioned theoretical perspectives; state allocation per department was the dependent measure. Because many of the measures used in the model exhibited relatively high intercorrelations (see appendix A) and the population being analyzed is relatively small ($N = 70$), special attention was paid to potentially problematic collinearity among the variables and to outliers that might be unduly influencing the parameters we report here. Parameter tolerances were assessed using a variance inflation factor (VIF) test derived from auxiliary regressions, condition indices were calculated, and the parameter variance was decomposed across each index. These tests indicated that collinearity was not a serious problem and that the data matrix was adequately conditioned for the analysis.[6] A residual analysis from the final model revealed four cases with absolute standardized residuals ranging from 2.23 to 2.79. Further analysis of these cases led us to conclude that they were not a threat to the estimates yielded by our final model.

Results

Our results are presented in four tables. Table 1 presents descriptive statistics on 14 key variables. Table 2 presents the OLS regression results from our final model. Table 3 displays key characteristics of the

TABLE 1

Means, Standard Deviations, and Descriptions of Variables

Description of Variable	Mean	SD	Minimum	Maximum	Variable Names
Dependent Variable (N = 70 departments)					
State Dollar Allocation	$1,693,338.90	$1,068,092.10	$135,812.00	$5,149,621.00	STATEDOL
Rational/Political Variables					
High quality–High centrality department	0.13	0.34	0	1	HICNTQAL
Number of graduate degrees awarded	20.02	22.01	0	106	GRADDEG
Number of undergrad degrees awarded	64.72	75.49	0	330	UGRADDEG
Total Student Credit Hours (SCH) generated	5310.86	4968.61	183	26227	TOTALSCH
Percent SCH at lower division	43.11	24.87	0	92.29	PCTLDSCH
Percent SCH at upper division	36.32	21.24	1	100	PCTUDSCH
Percent SCH at graduate level	20.57	18.67	0	88	PCTGRSCH
Grants & contracts / 1000	$1,184.34	$1,538.74	0	$8,714.52	GRANTS
Critical/Political Variables					
Percent faculty that are female	29.32	23.14	0	100	PCTFFAC
Percent faculty that are non-white	10.89	11.80	0	56	PCTMFAC
FTE lines to department faculty	1.04	0.43	0.15	2.53	FTERATIO
FTE student to faculty ratio	18.93	8.20	6	41	STUFACFT
Other Variables					
Average assistant professor salary/1000	$40.38	$7.39	$31.49	$65.94	SALASST
Average baccalaureate graduate salary/1000	$24.81	$5.52	$18.58	$43.75	GRADSAL

TABLE 2

OLS Slopes of Rational & Critical Variables on Allocation of State Dollars

Variable	Unstandardized Slope	Standard Error	Standardized Slope	VIF Score
HICNTQAL	628490	199266	0.20	1.29
GRADDEG	17469	4309	0.36	2.58
UGRADDEG	1368	1161	0.10	2.20
TOTALSCH	82	20	0.38	2.86
PCTUDSCH	−9646	4261	−0.19	2.35
PCTGRSCH	−20836	5870	−0.36	3.44
GRANTS	222	51	0.32	1.77
PCTFFAC	−1872	2909	−0.04	1.30
PCTMFAC	−5241	5457	−0.06	1.19
FTERATIO	−257922	154501	−0.10	1.29
STUFACFT	−43110	10587	−0.33	2.16
SALASST	42351	11661	0.29	2.13
GRADSAL	−29870	14730	−0.15	1.89
Intercept	1481990	550348		
Adjusted R^2	0.789			

campus's 9 high quality-high centrality departments, while Table 4 presents key characteristics of the bottom and top 5 departments (non-high quality–non-high centrality), sorted by state dollar allocations.

Discussion

The OLS regression results in Table 2 do not fully support either the rational/political model or the critical/political model. Instead, they suggest that a rational/political interpretation works for some institutional missions and a critical/political model for others. When market variables are introduced, both interpretations become more complex.

Generally, the positive slopes reported in Table 2 support the rational/political model. The relatively small standard errors associated with most of these slopes suggest, except in the case of undergraduate degrees, that there is not a great deal of variance among departments on these variables. On average, for every $1000 in grants a department brings in, a department receives an extra $222 in state dollars, all else equal. Given that approximately one-third of the annual operating budgets of the top public research universities are comprised of federal grants and contracts, these are substantial multipliers of departmental resources. The grants variable is rational/political in that the institution provides resource allocations to units that bring in external funds and, because most grants are competitive and peer reviewed, rewards research quality as well. High centrality-high quality[7] units enjoyed substantial rewards relative to

TABLE 3

Detailed Description of High Centrality–High Quality Departments (sorted by total state dollars received)

Department	Allocation	Grad. Degrees	UG Degrees	Asst. Prof. Salaries	Grants	% Female Faculty	% Min. Faculty	Total SCH	% SCH Upper Div	% SCH Lower Div	Students to Faculty	Faculty FTE Ratio
Pharmic. Sci.	$725,010	4	0	$50,032	$522,140	38	4	883	78	22	8	2.53
Hydrology	$1,293,609	20	9	$45,044	$2,769,984	7	21	1676	11	59	11	1.00
Art	$2,460,583	28	139	$31,486	$12,270	48	10	9186	34	11	17	1.14
Nursing	$2,693,291	35	85	$38,000	$1,867,531	100	8	3891	57	25	7	1.14
Psychology	$2,725,326	27	298	$36,000	$3,515,027	37	5	14515	42	8	26	0.96
Anthropology	$3,104,915	26	89	$35,667	$1,549,983	32	8	8779	37	15	18	0.98
Geological Sci.	$3,349,032	19	16	$36,680	$2,430,678	9	9	3808	14	23	9	0.97
Music	$3,705,301	31	46	$32,034	$160,079	30	12	9467	29	17	13	0.99
Astronomy	$4,313,180	37	8	$37,220	$8,714,516	0	0	5917	3	5	36	0.61

TABLE 4

Detailed Description of Non-High Centrality–Non-High Quality Departments (sorted by total state dollars received)

Department	Allocation	Grad. Degrees	UG Degrees	Asst. Prof. Salaries	Grants	% Female Faculty	% Min. Faculty	Total SCH	% SCH Upper Div	% SCH Lower Div	Students to Faculty	Faculty FTE Ratio
Bottom 5 in Terms of Dollar Allocation												
Judaic St.	$135,812	0	5	$40,370	$14,449	25	0	579	61	0	17	1.57
Med. Tech.	$263,764	0	3	$41,370	$0	100	0	327	100	0	7	1.00
Agric. Educ	$395,600	6	9	$44,510	$332,854	0	0	596	73	16	20	1.60
Statistics	$449,067	3	0	$38,000	$40,000	29	29	2986	12	14	35	1.12
Journalism	$558,058	16	66	$33,690	$91,716	29	10	2628	38	12	22	2.26
Top 5 in Terms of Dollar Allocation												
Chemistry	$3,234,861	31	21	$36,329	$5,570,259	8	8	14014	9	7	15	0.56
Physics	$3,308,575	21	25	$39,363	$2,948,320	2	12	8672	11	10	13	0.91
Math	$4,053,120	26	42	$39,378	$2,410,747	26	14	26227	10	6	22	1.25
ECE	$4,162,881	83	138	$46,325	$4,585,782	9	11	7901	41	32	13	0.93
English	$5,149,621	82	160	$33,071	$639,148	45	8	24681	30	8	19	0.98

other units. On average, those departments viewed by the faculty as central to the university mission and as being of high quality received an extra $628,490, all else equal. Although this is the largest effect we have, the reward for quality as a lump sum, unmultiplied, so it is in most cases a smaller allocation than those for grants, graduate student credit hours, and the like. The high centrality-high quality variable fits well with the rational/political model in that the departments seen as most important to the university's mission and the departments highest in quality receive greater state resource allocations than others. The small proportion of departments meeting the high quality-high centrality criteria (13% or 9 departments) clearly enjoy substantial financial benefits.

Workflow, at least in terms of SCH and degrees granted, to some extent supports a rational/political model. Net of other variables in the final model, departments receive $82 for every SCH generated, which, given the approximately 35,000 students at the institution in question, is a substantial multiplier. A premium is placed on lower division SCH, however. The $82 per SCH return is adjusted downward $9,646 for each 1% increase in upper-division SCH. Although the concentration of state dollars at the lower division runs counter to the rational/political conception of educational costs increasing as subject matter complexity increases as students move from the freshman year though graduate school, a rational/political case could certainly be made for allocating more institutional resources to departments with large numbers of lower division students. Such a rational/political interpretation would probably stress that: (1) lower-division students are least likely to persist, so concentrating resources on departments that teach them would serve the retention function, ultimately increasing graduation rates; (2) lower-division students take the most service courses, and the departments that provide these need the resources to do the job. Moreover, such a reward structure could also be seen as an incentive to streamline student course-taking within the department (i.e., for undergraduate majors and graduate students) while creating multiple lower-division course-taking opportunities for students from outside the department.

However, conferring institutional resources on departments that teach large numbers of undergraduate students does not seem to translate in a straightforward way to undergraduate degree completion. On average, departments receive $1,368 for every undergraduate degree completed, a relatively modest amount compared to resources for lower division SCH.

The institutional resource allocation pattern is focused more on lower-division SCH than on undergraduate degree completion. This suggests, although does not explain, systemic productivity problems if undergraduate degree completion is the paramount institutional mission. This resource allocation pattern does not fit perfectly either the rational/political model or the critical/political model, because neither model specifies what is expected to occur at the lower and upper divisions, but it is perhaps closer to the critical/political in that process (course-taking) is rewarded rather than product (degree completion). Emphasis on course taking benefits faculty and institution by generating SCH, rather than benefitting students or their parents, who would presumably value undergraduate degree completion most highly. However, this view is somewhat tempered by the earlier observation concerning incentives for efficiencies in upper-division SCH production that are undoubtedly linked to degree production.

Graduate school involves departmental allocation processes that are quite different than those of undergraduate school. Departments generating larger proportions of their SCH at the graduate level find they experience an offset of $20,836 per 1% increase in graduate SCH. The penalty to graduate SCH is, of course, relative to lower division SCH and due in large part to the nature of graduate education. Unlike undergraduate SCH, especially at the lower division, most graduate SCH are within the department or program in which the graduate student has matriculated. Nor is there a large SCH multiplier at the graduate level. The penalty to graduate SCH is partially offset by the return to graduate degrees, which is approximately 12 times greater than the return to undergraduate degrees. Graduate education is product centered, emphasizing completed degrees and the winning of federal grants and contracts. Undergraduate education, as we noted previously, is process-oriented: SCH is accumulated without degrees and grants.

There is a marked contrast between the resources allocated for completion of undergraduate and graduate degrees. While for every undergraduate degree granted a department receives $1,368, on average, departments receive $17,469 for every graduate degree they award. The rational/political model would expect the difference between undergraduate and graduate degrees because graduate education is more expensive due to greater subject matter complexity and greater student performance demands, which call for closer supervision of graduate students, all of which demands lower student to faculty ratios, which, of course, greatly increase costs. However, the rational/political model might not expect the very high dollar difference between graduate and undergraduate education that we found in Table 2. The differences in resource allocation to graduate and undergraduate degrees suggests that large public research universities have multiple and perhaps conflicting missions.

Like the rational/political model, the critical/political model predicts a difference between undergraduate and graduate degrees, but it would offer an alternative explanation. Rather than seeing graduate education as more expensive and the substantial dollar amount given for graduate students as functional, the critical/political model would see the high state dollar allocations of our model as excessive, unduly benefitting graduate students at the expense of undergraduates. The very large dollar difference between undergraduate and graduate degree completion may fit better with a critical/political than a rational/political interpretation.

Average assistant professor salaries, which we took to be a variable representing closeness of department to external markets, were consistent with market models. On average, for every $1000 increase is assistant professor salary, a department received an extra $42,351. Departments with high assistant professor salaries may be sufficiently resourced to put together more attractive packages to attract "hot" young professors, packages which include computers, laboratories and equipment, research assistants, and research support.

The negative slopes in Table 2 tell a somewhat different story, one that to some degree supports a critical/political model. We took the average salary of undergraduates on their first job after graduation, which we call student exit salary, as a variable indicating closeness of departments to external labor markets. Counterintuitive to market models, for every increase of $1000 to salary received by students on their first jobs, departments lost $29,870. When taken in conjunction with salary of assistant professor, another measure of closeness to external markets, institutions seem to be rewarding markets for faculty rather than markets for students. This suggests that institutions do not have a single market but multiple, and perhaps competing markets.

The variables associated with the critical/political model, percentage of female and minority faculty, confirm the expectation that departments with large numbers of women and large numbers of minorities receive less resources than other departments. This observation must be qualified by the large standard errors associated with these slopes which suggests that across departments there exists substantial variance in this association. For every increase in percentage of faculty who are female, departments lose $1,872. For every increase in the percentage of faculty who are racial minorities, departments lose $5,241. As noted above, there is great variation among departments on this measure.

The FTE ratio is the ratio of faculty headcount to the lines held by a department. This compares number of total faculty to the number of lines a department has, letting us see how the departments are rewarded when they use other than full-time, tenure-track faculty. For every line not held by a full-time, tenure-track faculty, departments lose $257,922. This finding suggests that departments that make heavy use of adjunct, part-time or off-track faculty or that are interdisciplinary, drawing heavily on faculty from other departments, are penalized in terms of state resource allocations.

The rational/political model does not specify use of nontrack faculty. The critical/political model sees the use of nontrack faculty as undercutting organizational stability, increasing the work load of track faculty, and generally substituting cheap labor for (appropriately) high salaried professional work. Contrarily, market theorists would interpret nontrack faculty, adjuncts and part-timers as a way of containing the costs and (appropriately) lowering resources to departments because the largest fixed cost within universities are faculty salaries.

The student FTE to faculty line ratio tells us that for every additional student per faculty line a department has, the department loses $43,110. In other words, the larger the faculty's student load, the fewer resources a department receives. As with the FTE faculty to line ratio, critical and market theorists would expect high faculty teaching loads to be unrewarded: these departments contain labor costs by hiring relatively few faculty relative to their student numbers and make heavy use of inexpensive adjunct faculty. However, critical/political theorists would condemn these practices as exploiting both faculty and students. Labor market theorists would see high student to faculty ratios as economies of scale. Conversely, rational/political models would expect departments with large student loads to receive more resources to facilitate workflow, an expectation our data do not confirm.

Overall, Table 2 suggests that there are at least two patterns for departmental resource allocation. One pattern generally fits the rational/political model in that departments that generate external funds, are central to SCH workflow, are high centrality-high quality, and confer a substantial number of graduate degrees receive high allocations of state dollars. The anomalies are concentration of resources at the lower division level and relatively low rewards for undergraduate degree completion, net of all other factors in the model. The other pattern is closer to the critical/political model, in which departments that have high percentages of female or minority faculty, use large numbers of off-track faculty, and have high student to faculty ratios receive lower state dollar allocations. However, there are a number of findings—such as loss of resources for upper-division student credit hours—that fit neither model and can be interpreted variously, depending on their relation to other variables.

Tables 3 and 4 allow us to see how the regression results in Table 2 play out in specific departments. Table 3 describes the 9 high quality-high centrality departments (recall that these departments enjoyed an average allocation premium of $624,490). Table 4 considers only those departments that failed to meet the campuses high quality-high centrality criteria and describes the 5 departments that receive the fewest state dollar allocations as well as the 5 departments that receive the most, allowing us to see the extremes.

Because Tables 3 and 4 are descriptive rather than analytic, we can use them only for illustrative purposes. We comment briefly on two departmental "Fat Cats" and then look at two departmental "Cash Cows" that are relatively unrewarded for their undergraduate labor.

The first of our Fat Cats is Astronomy, the last department ($4,313,180 in state dollars) in Table 3. Astronomy illustrates the (male) rational political model that rewards use of full time faculty, relatively large numbers of graduate students and high grant and contract dollars. Although the department teaches few upper division students and produces few undergraduate degrees, it nonetheless contributes to undergraduate education, producing large numbers of lower division student credit hours by teaching large service courses taken by students in the first few years. Indeed, this department may have a strategy of leveraging state dollars through large, nonfaculty intensive undergraduate courses while simultaneously leveraging federal grant and contract dollars through faculty intensive research and graduate education.

Our second Fat Cat is Chemistry, ($3,234,861 in state resource allocations), which appears fifth among the top 5 departments. Much like Astronomy, Chemistry capitalizes on the rational political model that rewards the use of full-time faculty, relatively large numbers of graduate students, and high grant and contract dollars. Chemistry generates a relatively large number of SCH with the vast majority (84%) at the lower division level. A notable difference between Chemistry and Astronomy is the former's low student to faculty ratio—something we would expect to see from a laboratory intensive department. While obviously larger than Astronomy, Chemistry generates about 1/3 less extramural grant and contract dollars. This amount is, however, still quite large and may also be used as leverage for state dollars for the department. It is quite clear from the large proportion of graduate degrees awarded and the significant extramural funding received that, despite the large proportion of lower division SCH produced, the Chemistry department prizes research and graduate programs.

Our first Cash Cow is Psychology ($2,725,326 in state allocations), the sixth department in Table 3. Psychology is in many ways a model, hard working but relatively underresourced department. From our model in Table 2 it is clear how Psychology's emphasis on undergraduate degrees (which also entails a commitment to less well rewarded upper division SCH) is a less lucrative enterprise than the graduate degree and lower-division SCH emphasis of the "Fat Cats" considered above.

Psychology also loses money for its relatively high percentage of female faculty. And of course, it loses due to its high student to faculty ratio. Its grants and contracts are at best a wash, considering its other liabilities. Yet, as we noted, Psychology produces more undergraduate degrees than any other displayed department.

Our second Cash Cow is Journalism ($558,058), the fifth of the bottom 5 departments. Journalism is very much like Psychology, except that is has fewer grants, and makes much greater use of adjunct faculty. Like Psychology, Journalism loses money for its relatively high percentage of female faculty and also for its relatively high percentage of minority faculty.

Our Fat Cats and Cash Cows suggest that some departments are better able than others to develop strategies for accruing resources. (Male) Fat Cat departments close to federal research markets (Astronomy, Chemistry) appear to be able to use grant funds to leverage state funds for more institutional resources. These departments serve lower division undergraduate students in service courses, for which they also receive relatively large amounts of institutional resources. Apparently these departments use their institutional resources to build their graduate programs, for which they then receive a premium for degrees completed. (Female) Cash Cows show that departments that award more undergraduate than graduate degrees (Psychology, Journalism), are far from federal grant and contract markets, and have high percentages of upper-division SCH are not as able to accrue institutional resources. Departments powerfully mediate individual academics' chances to receive institutional resources.

Conclusion

Looking at departments rather than at individuals or broad fields of study lets us see more clearly how state resources are associated with a range of variables used to explain institutional resource allocation. Looking at state resources rather than all resources was important because we were able to separate funds earmarked for research from moneys that traditionally have funded the core instructional mission (Gladieux, Hauptman, & Knapp, 1994). Departments were an appropriate unit of analysis because they are the organizational unit where the work of teaching, research and service are done, and they capture the way knowledge is organized.[8] We showed that departments powerfully mediate individuals' ability to accrue institutional resources. Funding for departments characterized by (male), full-time faculty, graduate degrees, grants, and contracts tend to be more highly resourced than departments characterized by (female) faculty, high (female) adjunct use, undergraduate SCH, particularly at the upper division level, and undergraduate degree granting.

None of the current models of resource allocation to departments—rational/political, critical/political, and market— fully explains resource allocation patterns. These models rely on underlying concepts of mission and market that may no longer be appropriate. Our conception of mission, first formulated in the last quarter of the nineteenth century, is rigid, unexamined, simplistic, and probably unsuited to the large, complex, bureaucratic organizations public research universities have become. Then, mission was conceived of as teaching, research, and service, the proportion of graduate to undergraduate students was very small, and most faculty were teachers who lacked PhDs and seldom engaged in research. Now, the numbers of graduate students have increased, most faculty have PhDs, and the time faculty devote to research has increased. (We never had measures for service, which also points to problems in our resource allocation models: how can we allocate rationally without performance measures?) Our study points to differentiation and conflict between undergraduate and graduate missions and points to why public research universities may have systemic productivity problems in undergraduate degree production.

Colleges and universities often try to clarify missions by focusing on markets. But markets are, like missions, multiple, complex and sometimes conflicting. As our study points out, the market(s) for newly degreed students may be very different from the market(s) for faculty, even when those faculty are assistant professors and are themselves relatively newly degreed. Instead, there are district and separate markets for faculty and for undergraduate students with newly completed degrees. Sometimes the markets for faculty are shaped by departmental proximity to federal research markets. There is not a linear market logic that dictates how state dollars are allocated to departments, so we cannot rely exclusively on market models.

We need to rethink our models of resource allocation as well as our ideas about mission and market. To do this, we will have to refine the categories we now use: for example, rather than teaching, we need to reconceptualize around lower-division teaching, upper-division teaching, and graduate education. We need to think about how teaching relates to degree completion and how that should be rewarded. Degree completion probably needs further refinement. For example, we did not distinguish between Master's degrees and PhDs, even though we suspect Master's degrees may be the Cash Cows of graduate education. When we think about research, we have to factor in how close departments are to federal research markets—in other words, are there substantial federal programs that can support research in the field—and if not, we have to decide whether to compensate such departments rather than penalize them. We have to conceptualize and operationalize service, something

K.P. -1107

APPENDIX A

Correlation Matrix ($N = 70$).

	STATEDOL	GRADDEG	UGRADDEG	GRADSAL	SALASST	GRANTS	PCTFFAC	PCTMFAC	TOTALSCH	PCTUDSCH	PCTGRSCH	HICNTQAL	STUFACFT
STATEDOL	1.0000												
GRADDEG	0.3770	1.0000											
UGRADDEG	0.2594	0.0753	1.0000										
GRADSAL2	−0.0334	−0.0496	−0.1557	1.0000									
SALASST2	−0.0201	−0.2609	0.0530	0.4878	1.0000								
GRANTS2	0.5698	0.2558	−0.1151	0.2238	0.0895	1.0000							
PCTFFAC	−0.1795	0.1431	0.0477	−0.1713	−0.2274	−0.2869	1.0000						
PCTMFAC	0.0031	0.0265	0.1280	0.0835	0.2264	−0.1462	−0.0215	1.0000					
TOTALSCH	0.6635	0.3051	0.4980	−0.2507	−0.2412	0.1241	−0.0224	0.1023	1.0000				
PCTUDSCH	−0.2243	−0.1689	0.3245	0.1491	0.3077	−0.2595	0.1849	−0.0138	−0.1703	1.0000			
PCTGRSCH	−0.2235	0.4046	−0.3582	0.2939	0.1415	0.0885	0.0345	0.0045	−0.4368	−0.2281			
HICNTQAL	0.3675	0.0914	0.0612	0.0744	−0.1238	0.3040	0.0686	−0.0715	0.0893	−0.0447	−0.0004	1.0000	
STUFACFT	0.0068	0.0019	0.4071	−0.4901	−0.1348	−0.0444	0.0300	−0.0065	0.3256	−0.0903	−0.4437	−0.1401	1.0000
FTERATIO	−0.2911	0.1456	−0.0571	0.0758	−0.0883	−0.1684	0.1105	−0.0319	−0.1766	0.1861	0.1556	0.0929	−0.0069

Notes

[1] The one exception to this was the Women's Studies Department, which was excluded because it was in its first year as a department and had only one faculty FTE allocated to it. We chose not to include departments within the medical school for several reasons. First, many public Research I universities, the category of institution we are examining, do not have medical schools. Second, medical schools have a disproportionate share of grant and contract dollars, given the great increases in funding by the National Institutes of Health in recent years, and we thought this would distort the effects of grants and contracts. Third, medical schools have an anomalous faculty-student ratio, often as low as one to one, which would again distort effects for the sample as a whole. Departments within the law school were also excluded, for no undergraduates are taught within this college.

[2] These scholars generally draw on that body of theory that focuses on race, class, and gender but often do not attend to day-to-day organizational politics, nor do they often employ quantitative data. For the political and quantitative dimension of the critical/political model, we draw heavily on feminist theorists who deal with issues of women and work, usually through quantitative data. (See Bellas, 1994, 1997; Bellas & Reskin, 1994; England, 1992; Pfeffer & Davis-Blake, 1987; Reskin, 1988; Reskin & Roos, 1990; Tolbert, 1986.)

[3] Our measure of state allocation per department is extremely conservative. We excluded overhead from federal grants and contracts from state dollars because the source was federal, not state. A substantial portion of the federal overheads, of course, accrue to the departments that win the grants and contracts, greatly increasing their departmental resources. The university central administration retains a portion of the overhead derived from each federal grant and spends the monies without restriction. The central administration's share of overheads may be reinvested in the very same departments or like departments that initially generated the grants, further concentrating resources on these departments.

[4] Academic labor market theories attempt to bring together human capital theories from economics and status attainment theories from sociology, but as Breneman(1988) points out, the two are as yet uneasily married, and neither inspects the linkages between academic labor markets and markets for scientists and professionals in the wider society (Breneman & Youn, 1988).

[5] In the Breneman and Youn (1988) volume, the most recent and comprehensive treatment of academic labor markets, gender is examined in only one article other than the single piece that addresses gender specifically. The Barbezat article, in that volume, focuses on and finds salary discrimination, as do all studies of salary differences between men and women in academe, but is cautiously optimistic about the reduction of discrimination over time (1968–1977), an interpretation not supported by more recent data (Bellas & Reskin, 1994). As Barbezat herself notes, her models examined only differences in relation of rewards to faculty characteristics, and as specified, cannot include gender differences in publication rates, nor institutional and departmental affiliations, all of which mediate the relation between faculty characteristics and rewards. Barbezat uses econometric models derived from the Oaxaca decompositions, which place a different set of constraints on the data than do the regression equations used by feminist researchers such as England (1992), Bellas (1994, 1997), and others. The implications of the differences between these approaches deserve further exploration.

[6] All VIF scores were below 5, and no condition index exceeded 30. These thresholds are shown to be viable in Belsey, Kuh, and Welsch (1980).

[7] Because quality and centrality were highly collinear in our preliminary models we combined these two variables in our final analysis.

[8] The role of centers and institutes in relation to departmental workload remains unclear. The few studies that separate centers and institutes from departments suggest they are much more closely associated with research than with teaching (Brooks, 1996).

References

Ashar, H. (1987). *Internal and external factors and their effect on a university's retrenchment decisions: Two theoretical perspectives.* Unpublished doctoral dissertation, University of Washington, Seattle.

Ashar, H., & Shapiro, J. W. (1990). Are retrenchment decisions rational? *Journal of Higher Education, 61,* 121–141.

Barbezat, D. (1988). Gender differences in the academic reward system. In D. W. Breneman & T. I. K. Youn (Eds.), *Academic labor markets and careers* (pp. 138–164). New York: The Falmer Press.

Bellas, M. L. (1994). Comparable worth in academia: The effects of faculty salaries on the sex composition and labor market conditions of academic disciplines. *American Sociology Review, 59,* 807–821.

Bellas, M. L. (1997). Disciplinary differences in faculty salaries: Does gender bias play a role? *Journal of Higher Education, 68,* 299–321.

Bellas, M. L., & Reskin, B. F. (1994, September-October). On comparable worth. *Academe,* 83–85.

Belsey, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of multicollinearity.* New York: Wiley.

Breneman, D. W. (1988). Research on academic labor markets: Past and future. In D. W. Breneman & T. I. K. Youn (Eds.), *Academic labor markets and careers* (pp. 200–208). New York: The Falmer Press.

Breneman, D. W., & Youn, T. I. K. (Eds). (1988). *Academic labor markets and careers.* Philadelphia: The Falmer Press.

Brinkman, P. (1990). Higher education cost function. In S. A. Hoenack & E. I. Collins (Eds.), *The economics of American universities.* Albany, NY: SUNY Press.

Brint, S. G. (1994). *In an age of experts: The changing role of professionals in politics and public life.* Princeton, NJ: Princeton University Press.

Brooks, H. (1996). Evolution of U.S. science policy. In Bruce Smith & Claude Barfield (Eds.), *Technology, R&D, and the economy* (pp, 15–48). Washington, DC: The Brookings Institution and the American Enterprise Institute for Public Policy Research.

Carnegie Foundation for the Advancement of Teaching. (1994). *A classification of institutions of higher education.* Princeton, NJ: Author.

Chickering, A., & Gamson, Z. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 39,* 3–7.

Dundar, H., & Lewis, D. R. (1994). Departmental productivity in American universities: Economies of scale and scope. *Economics of Education Review, 14*(2), 119–144.

El-Khawas, E. (1992). *Campus trends: 1992.* Washington, DC: American Council on Education.

England, P. (1992). *Comparable worth: Theories and evidence.* New York: Aldine de Gruyter.

Fairweather, J. S. (1996). *Faculty work and the public trust: Restoring the value of teaching and public service in American academic life.* Boston: Allyn and Bacon.

Gladieux, L. E., Hauptman, A. M., & Knapp, L. G. (1994). The federal government and higher education. In P. G. Altbach, R. O. Berdahl, & P. J. Gumport, *Higher education in American society* (3rd ed., pp. 125–154). Amherst, NY: Prometheus Books.

Gumport, P. J. (1993). The contested terrain of academic program reduction. *Journal of Higher Education, 64,* 283–311.

Gumport, P. J., & Pusser, B. (1996). Academic restructuring: Contemporary adaptation in higher education. In M. Peterson, D. Dill, & L. Mets (Eds.), *Planning strategies for the new millennium.* San Francisco: Jossey-Bass.

Guskin, A. (1994). Part II: Restructuring the role of faculty. *Change, 26,* 23–29.

Hackman, J. (1985). Power and centrality in the allocation of resources in colleges and universities. *Administrative Science Quarterly, 30,* 61–77.

Kerlin, S., & Dunlap, D. (1993). For richer, for poorer: Faculty morale in periods of austerity and retrenchment. *Journal of Higher Education, 64,* 348–377.

Konrad, A. M., & Pfeffer, J. (1990). Do you get what you deserve? Factors affecting the relationship between productivity and pay. *Administrative Science Quarterly, 35*(2), 258–85.

Layzell, D. T. (1996). Faculty workload and productivity: Recurrent issues and new imperatives. *Review of Higher Education, 19,* 267–281.

Lee, B. A., Leslie, D. W., & Olswang, S. G. (1987). Implications of comparable worth for academe. *Journal of Higher Education, 58,* 609–628.

Leslie, L. L., & Rhoades, G. (1995). Rising administrative costs: Seeking explanations. *Journal of Higher Education, 66,* 187–212.

Leslie, S. (1993). *The Cold War and American science: The military-industrial-academic complex at MIT and Stanford.* New York: Columbia.

Levin, H. M. (1991). Raising productivity in higher education. *Journal of Higher Education, 62,* 241–262.

McElrath, K. (1992). Gender, career disruption, and academic rewards. *Journal of Higher Education, 63,* 269–281.

Massy, W. F., Wilger, A. (1992, Winter). Productivity in postsecondary education: A new approach. *Educational Evaluation and Policy Analysis, 14,* 361–76.

Massy, W. F., & Zemsky, R. (1994). Faculty discretionary time: Departments and the `academic rachet.' *Journal of Higher Education, 65,* 1–22.

Melchiori, G. (1982). Smaller and better: The University of Michigan experience. *Research in Higher Education, 16,* 55–69.

Morgan, A. W. (1983). The new strategies: Roots, context and overview. In A. W. Morgan (Ed.), *Responding to new realities in funding.* San Francisco: Jossey-Bass.

Pfeffer, J., & Davis-Blake, A. (1987). The effect of the proportion of women on salaries: The case of college administrators. *Administrative Science Quarterly, 32,* 1–24.

Reskin, B. F. (1988). Bringing the men back in: Sex differentiation and the devaluing of women's work. *Gender and Society, 2,* 58–81.

Reskin, B. F., & Roos, P. A. (1990). *Job queues, gender queues.* Philadelphia: Temple University Press.

Slaughter, S. (1993). Retrenchment in the 1980s: The politics of prestige and gender. *Journal of Higher Education, 64,* 250–282.

Slaughter, S. (1997). Class, race, gender, and the construction of postsecondary curricula in the Unites States: Social movement, professionalization, and political economic theories of curricular change. *Journal of Curriculum Studies, 29*(1), 1–30.

Slaughter, S. (1998). Federal policy and supply-side institutional resource allocation at public research universities. *Review of Higher Education, 21,* 209–244.

Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies and the entrepreneurial university.* Baltimore: The Johns Hopkins University Press.

Spaeth, J. L. (1979). Vertical differentiation among occupations. *American Sociological Review, 49,* 746–762.

Stolzenberg, R. M. (1975). Occupations, labor markets, and the process of wage attainments. *American Sociological Review, 40,* 645–665.

Thurow, L. (1995). *The future of capitalism.* New York: W. Morrow & Co.

Tolbert, C. M., Patrick, M., & Beck, E. M. (1980). The structure of economic segmentation: A dual economy approach. *American Journal of Sociology, 80,* 1–10.

Tolbert, P. (1986). Organizations and inequality: Sources of earnings differences between male and female faculty. *Sociology of Education, 59,* 227–235.

Volk, C. (1995). *Assessing competing models of resource allocation at a public research university through multivariate analysis of state financing.* Unpublished doctoral dissertation, Center for the Study of Higher Education, University of Arizona, Tucson, AZ.

Youn, T. I. K. (1989). Effects of academic labor markets on academic career. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 5, pp. 134–154). New York: Agathon.

Dan Covell
Carol A. Barr

The Ties That Bind

Presidential Involvement with the Development of NCAA Division I Initial Eligibility Legislation

Introduction

> A college which is interested in producing professional athletes is not an educational institution. (Robert Hutchins, president, University of Chicago [Lawson & Ingham, 1980, p. 56])

> But America is different. Its universities are unique in their efforts to please many constituencies—prospective students, donors, legislators, the general public. The growth of intercollegiate sports aptly illustrates the strengths and weaknesses of a constituency-oriented system of higher education. With enthusiastic support from students, alumni, and even government officials, our colleges have developed athletic programs that have brought great satisfaction to thousands of athletes and millions of spectators. Few aspects of college life have done so much to win the favor of the public, build the loyalties of alumni, and engender lasting memories in the minds of student-athletes. (Derek Bok, president, Harvard University [Bok, 1985a, p. 124])

> College (sport) is what it is because the American public wants it so bad. . . . Now why the public wants it so much is a question for the public. Right? ("Temple Drake" [pseudonym], college sport insider [Morris, 1992, p. 93])

These statements identify an elemental conflict between academics and athletics that exists in American higher education; that is, the belief that the simultaneous institutional pursuits of rigorous academics and "big time" intercollegiate athletic programs are difficult, if not impossible, to reconcile. Many critics of American higher education note that our institutions are beset with contradictory and unrelated

Dan Covell is assistant professor of sport management, School of Business, Western New England College; Carol A. Barr is associate professor, Sport Management Program, University of Massachusetts at Amherst.

The Journal of Higher Education, Vol. 72, No. 4 (July/August 2001)
Copyright © 2001 by The Ohio State University

activities both academic and nonacademic in nature. The transformation of American higher education over the last century has led to criticism of academic activities—such as research funded by for-profit corporations—that often contribute little to students who fund the institution, and an unchecked academic balkanization on campuses has created a separation between undergraduate and graduate studies, arts and sciences, and liberal and professional learning that has meant confusion about the specific missions of specific institutions. Combine this with a current push for distance learning fueled by technological advances and the need to reach more diverse populations of students to maintain institutional and programmatic viability, and critics cite that it has become nearly impossible to define precisely what is meant by higher education.

This debate is made more complex when nonacademic components are also assessed in terms of their congruence with the mission of higher education. The adoption of the Cambridge/Oxford residential college model led to the encorporation of many nonacademic components within the traditional American higher education system, including intercollegiate athletics. This in part has led to the development of what Bok called the "constituency-oriented system of higher education," where schools use athletics and other nonacademic activities to foster a sense of community with students, alumni, and the general public. While the constituency-based system contains numerous potentially contradictory elements worthy of exploration, it is intercollegiate athletics that is often cited as a particularly aberrant aspect of American higher education, particularly at Division I National Collegiate Athletic Association (NCAA) institutions. The inevitable response of critics of the constituent system to this charge is, what do such activities have to do with the mission of higher education? The simultaneous pursuits of athletic success, related profits, and institutional academic integrity, say these critics, cannot be reconciled. To them, this is the glaring weakness in the constituent system. Supporters argue the strengths of the system, that the popular appeal of nonacademic activities are a vital complement to academic components and in keeping with the founding ethos of American higher education.

Efforts to wed the athletic and the academic attempt to deflect this criticism of the wedding of the athletic and the commercial that is inherent in the constituent system. According to Helman (1989), this ideal notion of intercollegiate athletics and the student-athlete is legitimized through eligibility rules, which provide "standards that tether commercial athletics to the educational purposes of higher education" (p. 237). If Division I programs are to meet the standards set by the NCAA that demand this tethering (see below), then the programs must be maintained and legitimized through such eligibility rules. The question that arises from this charge is to whom within the academy this responsibility of tethering will ultimately fall. It is in the realm of academic tethering that school presidents, the individuals who are seen to have ultimate control over all components of the campus, have moved to the fore. When first student-athletes and then faculty oversight groups proved unable to deal effectively with the problems associated with intercollegiate athletics and the demands of constituents, many school presidents saw it as their role as institutional CEOs, those managers who serve as the public face of the institution and the ultimate internal decision maker, to address these issues. Over time, certain groups of presidents have come to lead the associated public debate and NCAA organizational push for association-wide initial eligibility standards. Many other major concerns regarding Division I athletics—pay for play, controlling agent tampering, recruiting abuses by coaches, boosters and others, the recurring specter of gambling and point-shaving—have not elicited the same sort of demands for and responses of presidential leadership, because many presume that these are strictly "athletic" issues to be dealt with by professional athletic administrators.

In an attempt to understand the roles of presidents in maintaining congruence within the constituency-based American higher education system, this article provides a detailed chronology of presidential efforts to deal with the conflicts related to the tethering of academic mission to athletic pursuits through the development of NCAA initial eligibility academic legislation. Such legislation impacts recruiting and admissions, the ultimate sport product on the field and the court, and the charge to tether commercial athletics to the educational purposes of higher education and to preserve the viability of the intercollegiate athletic enterprise.

In response to criticisms that "big time" athletics has no place on campus and has no relation to institutional academic missions, the bylaws of the NCAA have been crafted to require that intercollegiate athletics be administered under an institution's academic rubric. The NCAA publishes annually the purposes of the association under Article 1 of its Constitution. The first stated purpose is, "To initiate, stimulate and improve intercollegiate athletics programs for student-athletes and to promote and develop educational leadership, physical fitness, athletics excellence and athletics participation as a recreational pursuit" (1999–2000 NCAA Division I Operating Manual, p. 1). Also included as stated purposes are, "To encourage its members to adopt eligibility rules to comply with satisfactory standards of scholarship, sportsmanship and amateurism," and "To legislate, through bylaws or by resolutions of a Convention, upon the subject of general concern to the members related to the administration" (p. 1). NCAA bylaws do not dictate whom schools may admit, as illustrated in Bylaw 2.5, "The Principle of Sound Academic Standards," which reads:

> Intercollegiate athletic programs shall be maintained as a vital component of the educational program, and student-athletes shall be an integral part of the student body. The admission, academic standing and academic progress of the student-athletes shall be consistent with the policies and standards adopted by the institution for the student body in general (p. 4).

An institution may admit any student, but the student may or may not be eligible to compete in intercollegiate athletics, depending on whether that student meets the initial academic eligibility criteria set by the NCAA membership. Division I schools must also recognize "the dual objective in its athletics program of serving both the university or college community (participants, student body, faculty-staff, alumni) and the general public (community, area, state, nation)" (p. 338), a verification of Bok's constituency-based assessment.

To host an NCAA Division I athletic program, therefore, a school must provide winning teams comprised of athletes who are also students, for the entertainment of those associated with the school and its constituents—students, faculty, community members, alumni, fans—to develop prominence at the national level and to strive for financial success. Such focus has been effective in developing the quality of the intercollegiate sport product and has led to the unprecedented levels of support from constituents and fans, as evidenced by the significant popular interest in such events as the NCAA Division I men's basketball tournament and the regular and bowl seasons of Division I-A football. And to win, to achieve national prominence, to develop the constituent interest to then attain the desired degree of financial self-sufficiency also required under the Division I philosophy statement, schools need skilled and proficient players. The subsequent pressure then comes from the efforts to attract, admit, and retain players on the basis of the athletic skills first and foremost, and those who possess such skills may or may not have any interest or abilities pertaining to academic pursuits.

Since the founding of the NCAA there have been thousands upon thousands of rule changes pertaining to the administration of intercollegiate athletics. Most focus on setting policy relating to and curbing abuses and perceived unfair institutional advantages in recruiting, financial aid, and issues amateurism. However, it is the idea of the student-athlete that is central to the entire intercollegiate athletic enterprise. The many associated supporting constituents of intercollegiate athletics compare favorably the college "game product," the contests as played by amateur student-athletes, to the professional game product, because to them college athletes play to represent their alma mater with pride, and play not for money but for the love of the game. However naive these beliefs may seem to many, they do persist. Thus Division I schools, charged with maintaining winning and profitable programs within stated academic boundaries, are also charged with maintaining this ideal notion of the student-athlete, the athlete who is also a student (what Sperber refers to as the "classic ideal" [1998, p. 368]), to maintain the popular and commercial appeal of its programs and to satisfy the demand inherent in the constituency. This, as author Willie Morris found when interviewing Temple Drake, is what the American public wants, and to maintain interest as prescribed in the constituency system and the NCAA bylaws, it is what must be provided.

Presidential Involvement with Intercollegiate Athletics and the NCAA

In the formative years of intercollegiate athletics, supporters extolled the virtue of athletics, especially football, in addressing the need to establish the virility of American males in an increasingly industrialized and mechanized society. Playing football, said MIT president Francis A. Walker, "demanded 'courage, coolness, steadiness of nerve, quickness of apprehension, resourcefulness, self-knowledge, self-reliance' and developed 'something akin to patriotism and public spirit'" (quoted in Telander, 1989, p. 32). This sort of advocacy portrayed intercollegiate athletics as a viable educational adjunct to academic classroom activities and a meaningful component of the constituency-oriented system.

However, as Rudolph noted, intercollegiate athletics quickly evolved to a system where during the early years of intercollegiate football, "Once the sport had been accepted, the games had to be won" (1990, p. 381). This emphasis was fueled by the interests of constituents—alumni and students—as Michigan president James G. Angell found when he and the Michigan faculty supported rules that would rein in the recruiting tactics by Wolverines football coach Fielding Yost. Heeding the protests of these constituents, with little concern for congruence between athletics and academic mission, the university's board of regents took control of the athletic program and subverted the control of Angell (Sack & Staurowsky, 1998). These actions signaled the clear establishment of the power of the constituents in the constituency system, and would be repeated throughout the century.

Some presidents were more successful in dealing with the problems of intercollegiate athletics. In 1905, when coaches and athletic administrators proved unable or unwilling to stem the spate of severe injuries and deaths in football, in part due to the interests of constituent groups and fans and of those coaches —such as Yale's Walter Camp—whose teams flourished under the old flying wedge systems, presidential involvement, as led by New York University's Henry McCracken, was integral in establishing the NCAA as a body for intercollegiate athletic reform (Yaeger, 1991). Columbia President Nicholas Butler stated that such efforts were "the first step in a general overhauling of the whole athletic situation in American colleges" (Smith, 1988).

With some exceptions, presidents demonstrated a limited role in the formation of NCAA initial academic eligibility policy. In the last two decades, as the expectation for harmony and congruence between athletics and academics and the demands of the constituency system have grown in light of a heightened perception of perceived ills within the management of intercollegiate athletics, presidents have gradually assumed an increased leadership role in this policy area, and have exhibited this influence through the mechanisms of the NCAA.

The NCAA and Initial Eligibility Legislation

The NCAA has struggled throughout its history to determine the appropriate combination of academic achievement and athletic prowess to be required of Division I student-athletes. It was resolved at the 1918 NCAA Convention "that in every college and university, the Department of Physical Training and Athletics should be recognized as a department of collegiate instruction, directly responsible to the college or university administration" (Falla, 1981, p. 55), and staff administrative committees with member school faculty and administrative personnel. Therefore, the expectation was put forth early on that schools are meant to control intercollegiate athletics, and athletics and student-athletes were meant to be integrated into the academic fabric of an institution.

The NCAA slowly assumed the role of arbiter for initial eligibility academic requirements, over the reservations of many member institutions. The primary goal of such a system would be to assure that all institutions would use the same minimum academic standards by which to assess prospective student-athletes in determining athletic eligibility and, in some cases, the appropriate financial aid award. Such an association-wide standard would serve to "level the playing field" for all schools, so that a student not admitted or eligible at school A would not later be deemed eligible at school B and suit up and compete for B against A. Former Harvard president Bok would refer to the need for broadly applied regulations because "many institutions will find it very hard to maintain reasonable academic standards if their competitors refuse to do likewise" (Bok, 1985b, p. 208). No school or president, no matter how willing they were to tether academics and athletics, was willing to do so alone and risk athletic ruin. In addition, such rules run counter to trends in American higher education, "which is remarkable for its decentralized arrangement characterized by institutional autonomy, voluntary association, and relatively little government regulation" (Thelin, 1994, p. 10), contributing to the overall sense of institutional incongruence and contradiction in the constituency-based system.

Major NCAA Academic Eligibility Actions

Supporting Rudolph's realization of the primacy of victory, Smith underscored the fact that to win games, "from an early period in American intercollegiate athletic history, there was pressure to bring in athletes with little regard for academic considerations" (1988, p. 175). Early conflicts developed over the participation of freshmen, graduate and nondegree "special" students, student residency requirements, and students transferring solely for athletic purposes. For roughly the first century of intercollegiate athletics, many colleges and universities maintained a separate admissions standard for athletes without abiding by the school's regular admissions policies. Thus, from the outset, the conflicts between academics and athletics were well established.

To deal with this conflict, some schools began to exhibit greater control over athletics, and took actions on their own to deal with academics, eventually usurping entirely formal control of athletic programs from students and alumni. The 1895 "Chicago Conference," attended by faculty representatives from the seven schools that later founded what would become the Big Ten Conference, was one such successful early effort to control and regulate the eligibility of student-athletes (Helman, 1989), as was the 1898 "Brown Conference," attended by constituents (students, faculty, and alumni) from nearly all schools of the present-day Ivy League, which in part sought to "weed out a student who has entered the university for athletic purposes only" (Smith, 1988, p. 140). Conference organizations would continue to serve as the main arbiter for such rules for many subsequent years, with varying degrees of effectiveness.

At the 1906 NCAA Convention, following the presidential impetus to form the organization, a policy concerning academic eligibility was adopted, citing, "No student shall represent a college or university in any intercollegiate game or contest who is not taking a full schedule of work as prescribed in the catalogue of the institution" (Falla, 1981, p. 144). It was left to individual student-athletes to voluntarily report any possible infractions. Smith (1988) points out that the group was formed on "the principle of individual (institutional) autonomy," and that "Home rule dominated the NCAA for the first half-century of its existence. The individual colleges agreed collectively to act individually" (p. 207). From these modest beginnings, the NCAA would eventually move toward stronger national initial eligibility legislation and enforcement procedures, but the process would take the better part of 50 years. With few exceptions, presidents would not assume a true leadership role within the association in this process until the 1980s.

Early Methods of Determining Initial Eligibility

By the early 1920s, most major conferences deemed freshmen ineligible for varsity level intercollegiate contests. Many dismantled these limitations in response to the manpower shortages that accompanied World War II and the Korean War, but some conferences reestablished them soon thereafter. Even though freshmen were ineligible to participate in varsity competition at certain schools, they were still eligible to receive athletically related aid (Mott, 1995). The NCAA finally approved full freshmen eligibility in all sports in 1972 (Sack & Staurowsky, 1998). In 1939 the NCAA established eligibility rules for National Collegiate championships. These rules were made more specific in 1946, stating that only those student-athletes admitted to their schools under the same admission standards as all other students would be eligible and that participants must be enrolled in a full course of study as defined by his school at the time of competition (Falla, 1981). Although these regulations were well-intentioned, their impact was questionable, as no enforcement mechanisms existed.

However, there were some attempts to institute reform on a broader and more meaningful scale at the institutional and conference level outside the NCAA. Individual schools had taken action to address perceived ills in athletics in the decade of the 1920s, with varying degrees of effectiveness. Butler at Columbia, Ernest M. Hopkins at Dartmouth, and Ernest H. Wilkins of Oberlin all spoke out against the expansion of intercollegiate sport. Said Wilkins, "Intercollegiate football is at the present time an enormously powerful force in the life of the nation," interfering "to an intolerable degree with the attainment of the purpose of the American college" (Betts, 1974, p. 347). Hopkins proposed a return to student coaches and participation limited to sophomores and juniors, while Butler advocated a new organization to replace the NCAA, an "Athletic League of Nations," where gate receipts would be abolished

(Betts, 1974, p. 348). In 1935 Southern Conference presidents, led by University of North Carolina president Frank Graham, recommended an end to remuneration for athletic ability and for all financial aid to be awarded by a faculty committee. Graham was to find little support for his plan, which though adopted in 1936, was abandoned in 1938. The regulations, said Graham, "cut too deeply into entrenched practice. . . . I am shocked to find that college presidents for this reason and that reason do not want to stand (in) back of the . . . regulations" (Sack & Staurowsky, 1998, p. 41). It was clear that the desires of constituents superseded concerns about tethering academics and athletics. It was also clear that one school acting alone in a reform effort was, as Graham found, doomed to fail.

The "Sanity Code"

As Graham found, a plan with nationwide scope, which would lift the political burden associated with threats to the college game product away from specific administrators and campuses, was needed to reconcile academics and athletics. The NCAA would gradually come to fill this role over the next four decades. The NCAA's first step in this process would come in 1946. At a "conference of conferences" in Chicago attended by faculty members and school presidents, attendees crafted a set of governing principles to deal with an acceleration of unsavory recruiting practices and athlete subsidies that had followed the conclusion of World War II. These were then sent to the NCAA membership for consideration (Helman, 1989). These principles came to be known as the "Sanity Code," due to a "prevailing belief that adherence to such principles [was] necessary to restore sanity to the conduct of intercollegiate athletics" (Falla, 1981, pp. 132–133). Sperber (1998) indicates this effort was an attempt by these established conferences to stem the new tide of competition for athletes and that the "Sanity Code" title was a revised epithet after the initial title, the "Purity Code," was derided as sanctimonious by scornful members of the press. However, the only penalty the NCAA could impose against those schools that failed to comply was expulsion from the association (Fleisher, Goff, & Tollison, 1992).

Many presidents also criticized the Code, due to concerns that the NCAA could only enforce the rule selectively, thereby leaving some schools at a competitive disadvantage as compared to others, and no president relished the idea of a losing program on his campus. In a speech after the convention, University of Notre Dame president John J. Cavanaugh criticized the reformers and the established conferences as "those who play with the question for public consumption, . . . true reform in athletics will not be accomplished by the mere publishing of noble, high-sounding codes which are often hypocritically evaded in actual practice" (Sperber, 1998, p. 186). Several school presidents indicated that their school would refuse to comply. University of Virginia president Colgate Darden, Jr., said, "While we may agree with the spirit of the plan, it is the belief of our board that it cannot be properly enforced" ("Purity Code Rejected," 1948, p. 28). Although Virginia and six other schools (Boston College, Maryland, the Citadel, Villanova, Virginia Military Institute, and Virginia Tech) were cited for violations of the code and expelled by the Compliance Committee, they were retained when a vote of the full NCAA membership failed to secure the necessary two-thirds majority for expulsion (Lawrence, 1987). Villanova president Francis McGuire expressed no regret for his school's alleged wrongdoings, but rather chose to criticize the Compliance Committee's efforts: "Do you mean to tell me," said McGuire, "that there are only seven schools in America which don't live up to the NCAA code?" (Sack & Staurowsky, 1998, p. 46). These schools vowed continued noncompliance, which led to a battle over the merits of the code at the 1951 Convention. Large schools and schools from the South lobbied successfully, by a vote of 130 to 60, to eviscerate the code (Lawrence, 1987). This action meant that no academic criteria would be utilized in determining student-athlete aid. In the case of the Sanity Code, many presidents demonstrated clearly that they were not immune to the pressure to attract skilled players to produce winning teams. In this instance, as in later cases, schools acted in a manner that was allowed them to assuage those pressures regarding constituent interests.

1.600 Rule

Following the demise of the Sanity Code, the establishment of formal athletic scholarships in the 1950s, as well as point-shaving scandals in college basketball and the test cheating by football players at West Point, renewed concerns about the tethering of athletics with institutional academic missions. Sack and Staurowsky (1998) point out that, "With an athletic scholarship system in place, it became absolutely imperative for the NCAA to establish a minimum academic level for awarding scholarships. To not do so would have fueled public cynicism that already surrounded professionalized college sport" (p. 96), for constituents demanded both game product excellence and preservation of the ideal notion of intercollegiate athletics. In response, the American Council on Education (described in its mission statement as the "nation's coordinating higher education association, with 1,800 members including accredited colleges and universities and other education-related organizations, . . . a forum for the discussion of major issues related to higher education," [Mission statement, 2000, p. 1]), commissioned a study of intercollegiate sport in October 1951, to make a recommendations to its 979 member schools. During ACE hearings for the report, chaired by John Hannah of Michigan State, Harvard football coach (and president of the National Football Coaches Association) Lloyd Jordan responded with little concern for congruence between athletics and academics and a particular understanding of the constituent-based system, "We must recognize that colleges are in the entertainment business. The only question is, How far should we go? But if we do it at all, shouldn't we do it well?" (as cited in Thelin, 1994, p. 108).

The final ACE report, presented in February 1952, called for more stringent eligibility rules, including admitting athletes based on the same standards as nonathletes, and a ban on postseason football bowl games. The nation's five higher education accrediting organizations, not the NCAA, would handle enforcement of these rules, with the loss of accreditation the penalty for failure to comply (Sperber, 1998). While Notre Dame's Cavanaugh and Yale president Whitney Griswold spoke out in favor of these reforms, Hannah, like colleagues Herman Lee Donovan at Kentucky and H. C. "Curly" Byrd at the University of Maryland, had fully embraced the constituent approach, and he used athletics to curry favor with politicians and businessmen within Michigan to improve the stature of his school and sought to protect Michigan State's competitive prospects. Sperber notes that Hannah and NCAA president Hugh Willett sought to forestall such change and to maintain the NCAA's control over intercollegiate athletics (Sperber, 1998). As a result, the ACE guidelines were coopted and watered down, and no meaningful academic eligibility reforms would be forthcoming that decade. Consequently, the provision of institution-based athletically related aid became entrenched, and that the NCAA would be the mechanism through which any initial academic eligibility rules would be established.

The bulk of other NCAA legislation over the next decade dealt with recruiting and compensation issues, including several proposals to severely restrict the scope of recruiting activities. The lack of significant presidential involvement with these issues bears out the role that presidents were expected to be less involved with these mostly "athletic" policy issues. However, the Association revisited academic standards in 1952, when an amendment to the association constitution required that all eligible student-athletes make normal progress toward a degree (Falla, 1981). But the critical task of determining normal progress was left to the discretion of each individual institution. Throughout the decade, several conferences, including the Pacific Coast and Big Ten took up the issue of admissions and academic eligibility standards. In 1952 the PCC presidents proposed unsuccessfully that student-athletes maintain a "C" average, while the Big Eight and Southwest conferences maintained GPA satisfactory progress requirements (Helman, 1989). The conference mechanism was one in which individual presidents could have more direct influence in an area over which they as the ultimate academic administrators were more empowered, and continued to serve as the main organizational element through which significant academic eligibility rules were established. However, in 1959 the NCAA did strengthen standards for championship competition, requiring that competing student-athletes be enrolled in a full course of study of no less than 12 semester or quarter hours (Falla, 1981).

At the 1962 Convention, the membership made the first serious attempt to impact student-athlete eligibility since the Sanity Code. Admissions was included with financial aid, eligibility, and recruiting as areas to be controlled under Association rules of conduct. Also that year, a study analyzed the academic records of 40,900 student-athletes at 80 member institutions during the 1963–64 school year (*1964–65 yearbook*, 1965). The findings of the study prompted the Long Range Planning and Academic and Testing Requirements committees to propose in 1964 and adopt in 1965 an expectancy table which established, as portrayed by NCAA President Robert Ray of the University of Iowa, "an academic floor for NCAA competition" ("Minimum academic," 1964, p. 4). According to committee member Laurence Woodruff, dean of students at the University of Kansas, the proposed expectancy table enabled colleges to "judge the probability that a student to whom we are granting an athletic scholarship will succeed academically in college" (Falla, 1981, p. 146).

At the 1965 Convention, the membership passed the association's first-ever minimum academic standards for the awarding of athletically related financial aid. The "1.600 rule" required that member institutions could not offer athletically related financial aid to student-athletes who did not have a predicted grade-point average of 1.600 (based on a maximum 4.000 scale) in the student-athletes' sixth, seventh, or eighth semester in high school (1964–65 yearbook, 1965). Standardized test scores on either the SAT or the ACT were also utilized in determining eligibility, which, according to longtime NCAA Executive Director Walter Byers, "provided an essential national comparative standard" (Byers, 1995, p. 158). Student-athletes were also required to maintain a 1.600 GPA while in college to remain eligible for athletically related aid.

That some opposition to the rule was voiced is not surprising, as schools saw the rule as a threat to their ability to meet constituent demands for game product excellence. But criticism and defiance came from unexpected sources as well, namely presidents of Ivy League schools. The schools in the conference were unwilling to submit to NCAA control over their admissions processes and academic autonomy, and chose to denounce the measure. Princeton president Robert Goheen charged that the rule "would appear to be the product of people more knowledgeable about athletics than of the life of the mind" (Jenkins, 1966, p. 30). Goheen's criticism had more to do with the NCAA's perceived intrusion into his and other Ivy institution's private affairs, renewing the debate over the appropriate extent of the NCAA's oversight. But no legislation was to be effective without a national scope, even though the University of Pennsylvania went as far as to flaunt the rule until faced with the possibility of being banned from the NCAA men's basketball postseason tournament. Goheen's stance notwithstanding, presidents still would not assume a significant role in setting the debate in this issue.

2.0 Rule

The booming popularity of intercollegiate athletics in the 1960s, fueled in large part by the emerging influence of television broadcasting and its associated revenues, led to a corresponding general lack of concern over the tethering of athletics to academics amongst the NCAA membership. To meet the demands of television, the media, and constituents, schools began to look even more toward attracting the best athletes possible to strengthen the game product, with even less regard for academic proclivities. This, in effect, spelled the end of 1.600, although attempts to weaken the 1.600 rule were defeated at almost every convention until 1973. Also, a 1972 Federal court case involving two University of California-Berkeley student-athletes, Isaac Curtis and Larry Brumsey, failed to diminish the scope of the rule. While a preliminary injunction was granted in favor of the plaintiffs, the court dismissed the case when the plaintiffs transferred to San Diego State University ("Cal suit," 1972).

In responses to other external concerns expressed by newly established constituent demands, presidential opinions to dismantle 1.600 were now beginning to be heard. Brubacher and Rudy (1976) point out that the Civil Rights Act of 1964 required that no college which received federal funds may practice racial discrimination and that subsequent "affirmative action" programs placed many schools under pressure to admit students of color, many of whom would be unable to compete under the standards of 1.600 (Helman, 1989). At the 1973 Convention, presidents

Robert Tierney of Queens College and Robert Behrman of City College of New York cited racial discrimination and loss of institutional autonomy as rationales to eradicate 1.600. Cornell president Frank Rhodes declared he would ignore the limitations of 1.600 and actively seek at-risk students in the inner cities in an effort to achieve racial balance on his campus (Byers, 1995). NCAA Secretary Treasurer Wilford Bailey, faculty athletics representative at Auburn University, later claimed that

> a major consideration in the decision to eliminate the 1.600 rule was the concern that retention of that requirement would result in reduced opportunities for minority students to participate in intercollegiate athletics, since an increasing number of colleges were adopting open-door admissions policies and Federally financed college aid programs for disadvantaged students were being expanded (Bailey, 1986, p. 3).

The fact that schools were now expected to provide educational opportunities to more groups, some of whom would be unduly penalized by restrictive initial eligibility rules, forced many presidents to reconsider the value of academic and athletic congruence. Issues of discrimination would increasingly frame the debate surrounding initial eligibility legislation, and many presidents were motivated by constituent pressure to remove 1.600 on these grounds.

Although NCAA President Alan Chapman defended 1.600 as "one of the few pieces of legislation we have on the books that does try to protect the student-athlete from exploitation for athletic purposes" (*Proceedings*, 1972, p. 167), the membership voted to repeal the rule in 1973, opting to replace it with a weaker version, the "2.0 rule." The new rule required that student-athletes who had graduated from high school with a 2.000 accumulative sixth, seventh, or eighth semester GPA (based on a maximum 4.000 scale), regardless of course content and test scores, were to be deemed eligible for participation and athletically related aid ("A history," 1980).

The retreat from 1.600, combined with the declaration of freshman eligibility in all Division I sports in 1972, led to many of the rules violations and abuses over the next decade which would serve to rekindle the debate in the coming years concerning the tethering of athletics and academics. In the face of an increasing public appetite for collegiate athletics and no real public concerns about these expansions (the point-shaving and cribbing scandals of the '50s were old news), presidential involvement to tether athletics and academics was not deemed necessary. Congruence between athletics and academics was not in demand, while excellence in the game product was. This, however, would soon change.

Proposal No. 48

> Many scholarship players are attending college for one purpose—to play college football and try to make it to the pros. They don't care about academics and neither do the coaches I've found that more and more college athletes are getting accepted to play college football who couldn't get into school before. (Dick Bestwick, Head football coach, University of Virginia, 1977 ["Opinions," 1977, p. 2])

The relaxing of standards resulting from the passage of the 2.0 Rule led to shifts in constituent perceptions of intercollegiate athletics. According to Thelin, "by 1978, the national press had once again rediscovered problems with coaching and recruiting excesses . . . despite the avoidance shown by some faculty senates and university presidents, academics and athletics would be intertwined in institutional and national policy debates" (1996, p. 178). Concerns regarding athlete exploitation, voiced by Alan Chapman after the demise of 1.600, would gradually become a significant issue in the formation of future initial eligibility legislation. The renewal of constituent concern in response to issues such as low graduation rates, star athletes who could not read or never attended class, and schools discarding untrained and uneducated athletes as soon as their athletic eligibility expired, would afford interested presidents the opportunity to assume a leadership role in the debate and to address these concerns through the NCAA for the reestablishment of a nationwide initial-eligibility standard. It would be they who would be identified as the group most able to deal with these problems in reestablishing the bonds between athletics and academics that had been loosened over the previous decades. As these bonds were destroyed, the result was damaging to the perception of the ideals notions of the student-athlete and intercollegiate athletics.

Despite Bestwick's observations (or rather, to bear them out), the membership continued to defeat attempts to enact more demanding initial eligibility requirements that might threaten the primacy of the on-field product. A "triple option" proposal, voted down by the membership at the 1978 Convention, provided for three alternatives for establishing initial eligibility: a high school GPA of 2.225, a minimum ACT of 17, or a minimum SAT of 750 ("New triple," 1978). Similar proposals were also voted down in 1979 and 1980 ("Are academic," 1980). E. John Larson of the University of Southern California wrote, "If the NCAA is to have any credibility regarding its posture on academic qualifications of prospective student athletes, the members must enact a bylaw at the January 1981 Convention that meets the scholarship goals of the constitution" ("Are academic," 1980, p. 3), but a proposal to increase the grade point requirement to 2.200 failed to reach the floor for a vote ("Governance approval," 1981).

In an ironic replay of actions from three decades previous, a more substantial call for reform came from once again from the American Council on Education. In 1982 presidents, via ACE, took a prominent role in promoting and establishing new, stricter initial eligibility rules. A proposal to confront the issue of initial eligibility requirements came forth from the ACE Ad Hoc Committee on the Problems of Major Intercollegiate Athletic Programs, chaired by Harvard president Bok. The committee cited the lack of control by presidents and trustees as one factor in the rise in intercollegiate athletic problems and called for increases in presidential involvement in athletic matters. The call for congruence was renewed.

The most substantive proposal put forth by the committee dealt with initial eligibility requirements. The proposal retained the concept of the 2.000 GPA Division I eligibility floor, but stipulated that the GPA would be computed from coursework

> in a core curriculum of at least 11 academic courses including at least three in English, two in mathematics, two in social science and two in natural or physical science (including as least one laboratory class, if offered by the high school) . . . as well as a 700 combined score on the SAT verbal and math sections or a 15 composite on the ACT (*Proceedings*, 1983, p. A-35)

According to Zingg (1983), the committee was aware of College Board findings that African-American students tested a full 100 points lower than whites on the SAT, and set the 700 standard,

> which essentially represented the midpoint for the lowest-scoring racial or ethnic group (blacks). The committee assumed that this floor would be acceptable. The committee assumed incorrectly The issue [was] not the establishment of an acceptable minimum test score—it [was] the use of standardized test scores at all. (p. 7)

The standardized test score component, eradicated with the removal of 1.600, had been resurrected and combined with a specification of what courses were considered to be college preparatory in nature. Bok and ACE leader Jack Peltason, former president of the University of Illinois, cited that the motivation for the proposal was to reestablish "the supremacy of academic values" in response to "the zeal to produce winning teams" (Greene, as cited in Funk, 1991, p. 108) that developed after decades of athletic expansions. Peltason noted that the 2.000 rule was "so meaningless that it is hypocritical to pretend that we are concerned first with education and only secondarily with athletics" (Wiedner, as cited in Helman, 1989, p. 215). Bok cited the dismantling of 1.600 as a prime factor in the increase in academic abuses, noting that USC admitted 330 special case athletes "without the participation of the admissions office" (1985a, p. 127). Most dropped out before their senior year (White, 1980).

Bok's role in this process, a departure from the stance taken by his Ivy colleagues regarding 1.600, speaks to the fact that any such national standard would probably be a competitive benefit to Harvard and the Ivy League schools, as other schools would be forced to adopt standards that were closer to those utilized in the Ivy League's admissions processes, thus leveling the national admissions playing field and potentially making Ivy teams more competitive nationally. In fact, the general stance of Ivy administrators and presidents recently has been to vote in favor of any legislation that increases the demands of academic standards, as it not only benefits the Ivy schools in terms of competition, but also because league members in general feel that promoting academic standards is an important philosophical stance to take.

Many of the committee members were in favor of a ban of freshman eligibility in football and men's basketball. Georgetown president Timothy Healy said, "(I) distrust the use of numbers, . . . what I'd like to see is the end of freshman eligibility. Let's give all students a year to get adjusted" (Fiske, 1983, p. 71), which would also serve to create a de facto national standard. George Hanford agreed with this approach, "which would get the questions of athletic eligibility for college freshman off the shoulders of the secondary school . . . and put it squarely where it belongs, in the hands of the colleges" (Hanford, 1985, p. 371). (The call for universal freshman ineligibility, which is resurrected periodically, has never been seriously considered to date. The recent increase in high-school seniors and collegiate undergraduates declaring eligibility for professional drafts makes such consideration even less likely, for such a ban would continue to limit the available number of athletes and therefore hurt the on-field sport product.) Such an action would also get eligibility off the shoulders of the College Entrance Examination Board. The formulation of initial eligibility requirements represented a compromise on the issue. The use of standardized test scores also added a controlled statistical standard uneasily influenced by local school grading practices and could provide a national measure that would be applied to all prospective student-athletes.

Drew University president Paul Hardin (who had resigned from the presidency at Southern Methodist University in 1974 after reporting to the NCAA rules violations in the SMU football program), captured the strongest motivation for the actions of the committee: "One or two presidents acting alone can effect a local situation but perhaps lose their jobs. If most of the presidents decided enough was enough, it could be done" (Gross, 1982, p. 7). As Bok and Hardin suggested, and as Graham had discovered in the 1930s, presidents needed to band together and become more visible, attending NCAA conventions [in fact, a then record number of presidents and CEOs—125—would attend the 1983 Convention ["125 CEOs," 1983], and 15 would speak to the merits of the ACE proposal [5 in opposition] ["Convention acts," 1983]). The NCAA would come to be the body through which interested presidents could and would act. (However, the NCAA's recent organizational restructuring, the move to a more federated legislative system and away from the one school, one vote convention system the association had used for the balance of the century, has, according to some, stunted the opportunity for a continued expansion of presidential influence. The current legislative process allows for the opinions of presidents to be expressed by the Board of Directors, comprised of presidents and chancellors. Legislative decisions are now made on a weekly basis rather than at the annual convention [Suggs, 1998]. This format has led many presidents, as well as other athletic administrators, to forgo attending the once important annual conventions, and has led many to feel estranged from the current decision-making process [Suggs, 1999a].)

L. Donald Shields, president of SMU and a member of ACE's Ad Hoc committee, spoke at the 1983 Convention on behalf of the proposal:

> It seems clear to many of us that in these days of increasing national concerns about the inadequate academic standards in our secondary schools and colleges that this legislation is not only appropriate but indeed is necessary to preserve the organizational integrity of the NCAA as well as the institutional integrity of our member institutions.... Our stewardship, our integrity as responsible leaders of institutions of higher education is at stake; ... we have to recognize that Proposal No. 48 represents reasonable, minimum academic qualifications for freshman eligibility; ... for us to leave this Convention without taking a significant action in this area of academic standards for student-athletes would be a travesty (*Proceedings*, 1983, p. 103).

Such constituent concerns were legitimate, for the graduation rates for athletes in high-profile sports were indeed poor. An NCAA study conducted from 1975 through 1980 discovered that only 42.9% of Division I-A football players graduated. A similar study conducted by *The Sporting News* found that 45% earned degrees, including a high of 100% in the Ivy League, to a low of 16.7% in the now defunct Southwest Conference (Morris, 1992). However, Allen Sack, sport sociologist and former Notre Dame football defensive end, and sociologist Harry Edwards criticized the plan as promoting standards that were too low. Sack (1986) later criticized the proposal as merely a weak attempt to assuage booster constituents to maintain interest in the ideal notions associated with intercollegiate athletics.

However, Edwards noted that "ironically, the most heated opposition to [proposal No. 48] came from black civil-rights leaders and black college presidents and educators—the very groups one might have expected to be most supportive of the action" (1983, p. 33), and he criticized the sports establishment, black communities, and the mass media for promoting an athletic career as the preeminent means for success for blacks. Indeed, Southern University president Jesse Stone, Jr., had criticized No. 48 because "a youngster knows that one way to move from the ghetto, if he's good enough, is to participate in college athletics and hopefully go on to the pros. . . . Without hopes and dreams, many of us would not have anything" (Farrell, 1985a, p. 33). That presidents at historically black colleges opposed the measure and the use of standardized test scores as racially discriminatory brought an ironic twist to the debate, citing the argument that weak students, many of whom were black, were being harmed by the same rules that were meant to eradicate their exploitation.

At the '83 convention, Joseph Johnson, president of Grambling State University, stated, "The ACE proposal blames the victim. . . . The committee's proposal . . . discriminates against student-athletes from low-income and minority group families by introducing arbitrary SAT and ACT cutoff scores as academic criteria for eligibility" (Proceedings, 1983, p. 103). His concerns were legitimate, because 49% of black male students failed to achieve at least a combined 700 on the SAT in 1981—as compared to 14% of whites and 27% of other minorities (Edwards, 1983). Johnson's statement indicates that not only would blacks be disproportionately harmed by No. 48, historically black schools would be disproportionately harmed in recruiting qualified student-athletes. However, Edward Fort, chancellor of North Carolina A & T, noted that reliance on standardized testing would hurt more than just African-Americans, citing a 1980 report on the Educational Testing Service (ETS, the administrator of the SAT) conducted by consumer activist Ralph Nader:

> The ETS and SAT scores discriminate not only against the rich and the minority of America but also between the rich and majority of Americans. That is, the members of the working and middles classes, black and white. . . . The more money a person's family makes, the higher that person tends to score. (*Proceedings*, 1983, p. 109)

These presidents were less concerned with congruence between academics and athletics and more concerned with how these rules would impact their schools' athletic programs and their ability to attract skilled athletes. Nonetheless, the measure was passed by the membership as an amendment to Article 5, Section 1-(j) of the NCAA Bylaws and would take effect on August 1, 1986, with the incoming class of 1990. A proposal that called for the awarding of athletically related aid for those freshmen who met either the GPA or test score (termed "partial-qualifiers") was also accepted.

In the pursuit of reestablishing initial eligibility guidelines and with the passage of No. 48, presidents had finally established themselves as a group able to craft, to lobby for, and to pass NCAA legislation as never before, in large part due to their ability to recognize strength through numbers and learn from the fate of predecessors like Frank Graham. The motivation and ability to act as a group was also more substantial than ever before, because the issue of academic standards was one upon which presidents, as ultimately responsible for preserving the academic mission of higher education, could stand without fear of criticism that they were not professional intercollegiate athletic managers and were treading in areas where they had little expertise. If a school president could not stand up to reinforce the importance of the academic component of intercollegiate athletics, who, with any real legitimacy, could? The presidents had found an issue that, in general, they as a group could agree was important; they could, for the most part, agree on a recommended course of action and could recognize that the majority of the media and the general public approved of their actions. Few of the many other problematic issues relating to intercollegiate athletics could garner the same unanimity of purpose and action.

No. 48 Fallout

The passage of No. 48, while serving to rally and unite many presidents to action, created a new set of problems for supporting presidents and the NCAA. Regardless of the fact that No. 48 was the NCAA's most significant attempt to tether athletics and academics to enable programs to meet Division I guidelines, the NCAA could not dodge the repeated criticisms that these newly established guidelines were disproportionately penalizing African Americans. In September 1984, data from an NCAA-sponsored study performed by Advanced Technology, Inc., of 16,000 male and female student-athletes at 75% of Division I schools on the impact of No. 48, including the relationship between graduation rates and standardized test scores, showed that only 18% of black male student-athletes admitted in 1977 and 1982 would have qualified for freshman participation, while 27% of black females, 57% of white males, and 60% of white females would have qualified (Farrell, 1984). The study also found that if either of the criteria (test scores or GPA) were utilized independently, 51% of black males would qualify, compared to 52% of black females, 91% of white males, and 93% of white females. These findings, coupled with the claims that the standards were racially biased, would prompt presidents and the membership to enact some revisions.

In October 1984, the NCAA Council recommended to the newly established Presidents Commission to delay for two years the implementation of 5-1-(j) ("NCAA to seek," 1984).
(At the 1984 NCAA Convention, in part due to the successes in the promotion and passage of Proposal 48 [Bok, 1983], ACE proposed that the membership act to formally include presidents in the decision-making process of the organization. The membership approved the establishment of a 44-member "Presidents Commission," but rejected an additional ACE proposal intended to empower the newly established commission to suspend the association's rules and impose regulations of their own in an effort to address more effectively the perceived ills affecting intercollegiate athletics [Vance, 1984]. The ACE measure was attacked as putting too much power into the hands of too few college presidents, a notion that was unsettling to many athletic administrators as well as to some presidents. However, those president who sought to impact NCAA policy now had a formalized channel through which they could act.

Presidents also reestablished an active voice in calls for reform in the 1990s through the advent of the Knight Foundation Commission on Intercollegiate Athletics. The Commission, sponsored by the Knight Foundation, was initiated by Foundation head Creed Black in 1989, a former Kentucky newspaper publisher whose paper ran stories exposing athletic abuses at the University of Kentucky. Black, who had his life threatened by irate UK fans because of the articles, organized the Commission to review problems in intercollegiate athletics. The committee brought together presidents, CEOs, and athletic directors from Division I schools [DiBiaggio, 1995]. Among its many recommendations in its three published reports [the committee was dissolved in 1993], it was suggested that presidents could effect these changes in part by controlling the NCAA through attendance at the annual Convention and overseeing athletic affairs more closely [Knight Foundation, 1991].)

In 1985 the NCAA's Special Committee on Academic Research (which included presidents Fort and Johnson ["Modifications in," 1985]) considered eliminating the test score requirement ("New academic," 1984). The Committee recommended that the minimum ACT score be lowered from 15 to 13. Committee chair Wilford Bailey justified this change:

> The purpose of the test score . . . is to assist in identifying students who are an academic risk. . . . Changing the ACT minimum requirement from 15 to 13 would place the cutoff score for each test at approximately the same percentile of students taking each test nationally; . . . the proposed modifications eliminate the arbitrary test-score requirement and reduce the disproportionate impact on black students without changing appreciably the percentage of students graduating ("Two changes," 1985, p. 12).

This change meant that, once factored into the eligibility index, approximately 24% of black males would now qualify as eligible as freshmen, utilizing the 13 score, while 38% would qualify utilizing the 700 SAT score. It was also projected that graduation rates for black males would increase to 68% ("Special committee's," 1985). In response to these findings, the NCAA Council and Presidents Commission agreed in November 1985 to sponsor legislation at the 1986 Convention that would amend 5-1-(j) to incorporate a sliding scale for determining eligibility, allowing a lower GPA to be offset by a higher standardized test score. Minimum thresholds for each standard were lowered to 1.800 (from 2.000) and 660 and 13 (from 700 and 15) (Farrell, 1985b). This proposal, No. 16, would amend the entrance standards for freshman student-athletes entering in the 1986–87 academic year and called for the standards to be adjusted to 1.900, 680, and 14, respectively, for the following year.

In response to criticisms that this new proposal represented a retreat from the standards set by No. 48, Indiana University president John Ryan, chair of the Presidents Commission, said, "I do not in any way believe that it represents a watering down of the original intent of bylaw 5-1-(j)" ("NCAA debates," 1986, p. 34). Said Prairie View A&M University president Percy Pierre: "It is not perfect, but it is much less flawed. . . . It is only 50-percent worse for blacks than whites, while (No. 48) is 100-percent worse for blacks than whites" (Wieberg, 1986, p. 6C). Pierre's response indicated that attempts were made to address some of the primary concerns of presidents of historically black colleges and universities. At the 1986 Convention, Wilford Bailey spoke in support of No. 16, and addressed this concern by noting that

> in this period of transition [no. 16] addresses an aspect of fairness that I think has not been specifically identified. That is a fairness to individual students who may fall slightly below the standard in this transition period, one or the other of these requirements, and who could be accommodated here without reducing in any appreciable way the academic validity of the requirements as it is stated (*Proceedings*, 1986, p. 78).

No. 16 was approved by a vote of 209 to 95, with four abstentions.

Proposition 42

As is common with most if not all new NCAA legislation, the establishment of initial eligibility requirements led some to believe that the new rules produced, along with gains in academic integrity, the opportunity for concocting innovative methods to circumvent both the letter and spirit of these rules. In response to this concern, a proposal was brought forth at the 1989 Convention to close what was considered to be significant loopholes in the initial eligibility requirements. In this case it was the option currently allowed under Bylaw 5-1-(j) where "partial qualifiers," those prospective student-athletes who met either (but not both) of the testing or GPA standards, were permitted to receive athletically related financial aid (at the cost of a lost year of athletic eligibility). Critics saw this as a way to still allow schools who wished to admit and fund unqualified prospects, lessening the intended impact of No. 48. The proposal that sought to change this, listed on the '89 convention business agenda as Proposal No. 42, would eliminate the partial qualifier and restrict athletically related aid to qualifiers (those who met both standards) only. A 1988 NCAA survey showed that at 282 Division I schools, 4.5% of enrolled freshman student-athletes were partial qualifiers (65% of whom were black), while 3.4% were nonqualifiers ("Results of," 1988).

The call to eliminate this perceived loophole came from the Southeastern Conference, which had opted to eliminate the awarding of funding to partial qualifiers by 1993 (*Proceedings*, 1989) in response to an academic scandal at the University of Georgia. In 1986, Le Roy Ervin, a University of Georgia academic affairs administrator who ordered remedial studies teacher Jan Kemp to change athletes' grades, said, "I know for a fact that these kids would not be here if it were not for their utility to the institution. There is no real academic reason for their being here other than to be utilized to produce income" (Nack, as cited in Funk, 1991, p. 76). In the wake of the scandal, Georgia president Fred Davison moved in 1986 to no longer permit the admission of nonqualifying recruits at his school in an effort to reestablish the ideal notion of the student-athlete. Georgia coaches and athletic officials, fearing future recruiting and competitive disadvantages that would result from this restriction, and the negative reaction of constituent groups that would follow if Georgia teams floundered, convinced other SEC schools to ban nonqualifiers conference wide in 1988 (Oberlander, 1989). Many other conference commissioners publicly praised the rule, but stopped short of endorsing it for their conferences, not willing to give up the option to recruit prospects that might help their schools' programs. Said Big Ten commissioner Wayne Duke, "I think it is a remarkable and courageous step. It will be interesting to see how it evolves" (Barnhart, 1989).

After adopting the Georgia rule, the SEC then sought to impress this approach upon the rest of the NCAA membership. For it stood to reason that if other conferences continued to allow athletically related aid to partial qualifiers, SEC schools as a whole would experience recruiting and competitive disadvantages. In response, the SEC announced in August 1988 that the conference would introduce legislation at the 1989 Convention to formally ban athletically related aid to partial qualifiers by 1993—the same rule as the SEC had enacted earlier that same year ("SEC planning," 1988). Douglas Hobbs of UCLA, chair of the Academic Requirements Committee, advocated for the defeat of the SEC's proposal at the Convention, advising not "to tinker with (No.) 48 . . . until we have experienced four, preferably five years, of admitting classes" (*Proceedings*, 1989, p. 247). In fact, NCAA director of research Ursula Walsh remarked in 1987 that definitive results of the impact of No. 48 would not be known for a decade (Lilley, 1987). The membership agreed, and the proposal was defeated on a close vote, 159 to 151, with four abstentions.

But No. 42 was not dead. The next morning (January 11), a reconsideration of the proposal was initiated by D. Alan Williams of the University of Virginia. Said Williams:

> We, as with several other institutions, responded to the request of the Presidents Commission not to make any changes in Bylaw 5-1-(j); but upon further reflection and getting a better reading to the degree to which the Presidents Committee really discussed this, we are prepared to vote in favor of Proposition 42 (*Proceedings*, 1989, p. 279).

Regardless of the calls by the Presidents Commission and others to hold off on alterations on initial eligibility, overnight lobbying contributed to the change in heart Williams expressed. The delegates from the SEC "did some of their more intense backroom lobbying" (Oberlander, 1989, p. A36) following Tuesday's defeat, especially of ACC delegates like Williams. The next day's reconsideration debate lasted twenty minutes ("Vote confused," 1989). On a second vote, Proposal No. 42 was passed by a slim margin, 163 to 154, with two abstentions. "You certainly can't say this is a mandate from the membership," said NCAA Executive Director Richard Schultz (Rhoden, 1989a, p. A25). The passage assured that, beginning in August 1990, no schools could provide athletically related aid to partial-qualifiers (*Proceedings*, 1989).

However, the criticism of the disparate impact of initial eligibility on minorities intensified as a result of this change. The National Center for Fair and Open Testing publicly challenged the NCAA to justify the use of standardized test scores, or remove them ("Testing service," 1989). Some convention delegates claimed they had been confused by the question and said they voted in favor of eliminating the partial qualifier when they thought they were voting to keep it. The vote was taken by roll call, with each school responsible to punch out their vote on computer cards. Said Jim Marchiony, Director of Communications for the NCAA: "It's as easy as in all-star voting, even easier. You do not have to be a rocket scientist" ("Vote confused," 1989, p. 7C). Kirkland Hall, AD at University of Maryland-Eastern Shore, said he was under the impression that the new rule would allow nonqualifiers to receive institutional aid (Moore, 1989). Said Hall, "I don't think it was explained properly" (Bannon, 1989, p. 1C). The day after passage, SEC commissioner Dr. Harvey Schiller erroneously reported that No. 42 "doesn't preclude (nonqualifier) admission or financial aid. The only thing it precludes is the awarding of athletically related aid" (Wieberg, 1989a, p. 1C). As passed, No. 42 would allow recruited nonqualifiers assistance only from state and federal grants, and they would receive no institutional aid whatsoever. Acting ACE president Robert Atwell said the rule was "just plain wrong if needy students aren't able to get any kind of institutional aid" (Oberlander, 1989, p. A1).

While most presidents did not support the change, the most vociferous detractors of the rule were two African American head coaches of prominent Division I men's basketball programs: John Chaney of Temple University and John Thompson of Georgetown University. The day after passage, Chaney, who had five ineligible freshmen in his program, tore into the NCAA: "I've no confidence at all in that racist organization making a decision on behalf of black youngsters. They've gone far beyond what I figured anybody who considered themselves interested in education would go," and suggested that black schools could "leave the damned NCAA and form a league of our own" (Wieberg, 1989a, p. 1C). Chaney also blamed "racist presidents" for conspiring to pass the legislation ("Thompson to," 1989, p. 69).

Two days after the passage of No. 42, Thompson announced that he would not be on the bench for third-ranked Georgetown's game against Boston College on January 14. Thompson devised the protest to draw attention to the "tremendous tragedy" that No. 42 would cause, and that he would not be on the bench during a game "until I am satisfied that something has been done to provide these student-athletes with appropriate opportunity and hope for access to a college education" (Weaver, 1989, p. 45). Thompson stated that he would petition the NCAA and seek to enlist other coaches (Weaver, 1989). Georgetown president (and Presidents Commission member) Rev. Timothy Healy backed Thompson, and called No. 48 "unwarranted and unacceptable an intrusion in the college's business" (Weaver, 1989, p. 45). Thompson's protest did not end following the Hoyas' 86–60 victory over BC (Douchant, 1989a). He did not accompany his team for its game at Providence College on January 18 (Schuster, 1989).

Constituent reaction, combined with Thompson's public actions and comments, were starting to make an impact. *USA Today* reported that of 65 schools surveyed, 16 would consider changing their "yes" vote. In response to racism claims, SEC commissioner Schiller said, "If any part of NCAA rules (are) diminishing the opportunities for minorities, then it ought to be examined and we ought to eliminate them" ("Testing service," 1989, p. 1). Presidents Commission chair, Dr. Martin Massengale, chancellor of the University of Nebraska, admitted that No. 42 would be raised for discussion at an upcoming Commission meeting, and that the "Commission might ask for reconsideration. . . . I hope it can be worked out without a special convention" (Schuster, 1989, p. 1C). The next day, Thompson, Healy, and Georgetown AD Francis Rienzo met with Schiller, Schultz, Massengale, and NCAA Council president Albert Witte (professor of law at the University of Alabama) to discuss No. 42. Following the summit, Massengale and Witte stated that they would recommend postponing the implementation of No. 42 in the form of proposed legislation at the 1990 Convention ("'90 Convention," 1989). Thompson then announced,

> I've decided to return to coaching my team based on the fact that I think there's a sincere effort on the part of the people in the N.C.A.A., the two presidents, Dr. Massengale and Witte, to make a sincere commitment to take this thing back before the convention and re-evaluate what has happened. . . . No one had to be beaten over the head. It was a sensible, intelligent discussion. (Rhoden, 1989b, p. A23)

By February, nearly 40% of those who had voted in favor of No. 42 now expressed opposition to the measure ("Opinions shifting," 1989). The efforts to maintain congruence through the closing of the perceived loophole were now viewed as incongruent with the mission of education, as perceived by many, which was to provide educational access to minority students. In October, Schultz said at a meeting of the NCAA Council: "There are a lot of divergent feelings out there. . . . I don't think there's any doubt it's going to be revised some way. . . . I know there's going to be a proposal to delay it. And I'd be surprised if there wasn't a proposal to eliminate it" (Wieberg, 1989b, p. 5C). Schultz noted that the Presidents Commission had already drafted a proposal for the 1990 Convention that would allow partial qualifiers to receive need-based financial aid (Wieberg, 1989b). Schultz, however, did not waiver in his support of the initial eligibility requirements: "You hear a lot about the fact that 90 percent of nonqualifiers are black, but that's a misleading number because 84 percent of black athletes have qualified under (No.) 48" (Ames, 1989, p. 68). Schultz would also later say that had Thompson not undertaken his public protest, "you would have had some coaches make some negative comments about it. It would have been in the paper for two or three days, and that's all you would have heard about it" ("But for," 1989, p. 16). Personal and constituent vested interests were a factor in the actions of Thompson and others as well, for No. 42 would have hindered those schools that most heavily depended on recruiting those athletes most directly impacted by the rule. Much of the argument against No. 42 was presented in terms of protecting academic access for needy minority students, which was recognized as a legitimate constituent concern and congruent with the mission of higher education, even though its impact was focused in athletics, a clearly nonacademic component of the American higher education system.

Proposal No. 26

Following the furor over No. 42, several proposals related to No. 42 were considered by the membership at the 1990 Convention. The Presidents Commission did indeed sponsor a measure to restructure allowance of aid for partial qualifiers. Part of the proposal, No. 26, set forth to amend what was now Bylaw 14.3.2 so that

> an entering freshman with no previous college attendance who enrolls in a Division I institution and who is a partial qualifier . . . may receive institutional financial aid based on financial need only, consistent with institutional and conference regulations, but may not practice or compete during the first academic year in residence (1990 Proceedings, 1990, p. A-26).

Much lobbying in favor of approving No. 26 occurred previous to the Convention. In December the Black Coaches Association (BCA), an organization of 2,500 college and high-school coaches, wrote letters to Division I athletic directors advocating for the passage of No. 26. Said BCA executive director Rudy Washington, then an assistant men's basketball coach at Iowa, "Leave admissions requests up to the academic institutions, then let us get back to the business of being educators" (Robbins, 1990, p. 13).

There was some debate on the convention floor regarding particulars of other proposed amendments, which dealt with how to define recruits and official visits prior to the vote on No. 26, but the decision had been made before the delegates arrived. When presenting No. 26, UCLA's Young, a member of the Presidents Commission, stated that the institutional financial aid could not come from the athletic department, and thus the proposal "would retain the academic incentives that are inherent in the Association's initial legislative eligibility without the devastating financial side effects that the 1989 Proposal No. 42 would create" (*Proceedings*, 1990, pp. 196–197). Little other floor debate followed, although Penn State football coach Joe Paterno said: "I can tell you 48 is working. Board scores are coming in better, and grade-point averages are coming in better. If we rescind 42, we're saying to the kid, Ok, you really don't have to study, because there is back door you can come through' " (Missanelli, 1990b, p. 4-F).While Paterno's comments indicated that No. 48 was providing a vital measure of congruence, constituent demands indicated that access for needy minority students also should be congruent with institutional missions. The membership agreed, and the measure was approved easily, 258 to 66, with one abstention (*Proceedings*, 1990).

Proposal 16

Following No. 26, the legislation surrounding initial eligibility changed little from the original intent associated with No. 48. Many studies were published in the years after No. 26, and while some detected a relationship between initial eligibility standards and increased academic performance, others did not. Research performed by the NCAA failed to show a positive correlation between initial eligibility standards and improved academic performance. A five-year longitudinal study of over 180 Division I schools completed in 1991 showed that over the period, the percentage of overall scholarships awarded to student-athletes who lost eligibility due to Proposal No. 48 grew from 4.5% in 1997–88 to 5.6% in 1990–91. The percentage of these student-athletes who were black rose over that same period from 65% to 68.6% (Wieberg, 1991). The NCAA study also showed a similar relationship to graduation rates for both standardized test scores and high school GPA ("Grade-point," 1991).

Following the retreat from No. 42, the Presidents Commission would again consider raising initial eligibility standards. The reassessment in part can be attributed to the fact that although the impact of initial eligibility standards was inconclusive, constituent reactions to the measures to tether athletics and academics were generally favorable. Experience also showed that only national standards would ultimately be effective. At the 1991 Convention, the Presidents Commission and the NCAA

Council proposed a resolution that the Academic Requirements Committee "review the recent research data and to recommend legislation to strengthen the current NCAA requirements for both initial eligibility and continuing eligibility" (*Proceedings*, 1991, p. A-113). Said R. Gerald Turner, president of the University of Mississippi, "I strongly urge your 'yes' vote on this important message, both to the Convention and higher education, and to the American public" (*Proceedings*, 1991, p. 355). The proposal was adopted by all divisions of the membership.

In June 1991, the Commission endorsed a proposal, effective August 1, 1995, that would raise the number of core courses required of freshmen from 11 to 13, and a sliding scale correlating high-school GPA with standardized test scores. Currently, incoming freshmen were required to have successfully completed 11 core courses (3 years of English, 2 years of mathematics, 2 years of social science, 2 years of natural or physical science) (1992 Proceedings, 1992). A freshman with a GPA of 2.000 would now need a combined SAT score of 900; a GPA of 2.250 required a combined 800; a GPA of 2.500 required a score of 700. Said Schultz, "The presidents are on the right track. When it's passed in January—and I feel comfortable it will be—it will be a historic day" (Davidson, 1991, p. E15). Commission Division III subcommittee chair David Warren of Ohio Wesleyan University captured the essence of the Commissions' motivations for action: "The public is concerned about the exploitation of athletes, about bringing them in as Hessians and using them as fodder. I think both the public and the Presidents Commission are determined to alter both the perception and reality of this" (Asher, 1991, p. B13). Marquette University president Rev. Albert DiUlio echoed this, stating, "There has been extraordinarily strong support for making initial eligibility requirements more restrictive" (Asher, 1991, p. B13). Results were important; the public's perceptions were equally so. DiUlio also noted that the Commission was acting in consort with the recommendations of the Knight Foundation Commission on Intercollegiate Athletics (Asher, 1991). The two African American members of the Commission, North Carolina Central chancellor Tyronza Richmond and Mississippi Valley State president William Sutton, continued to express the opinion of historically black colleges and universities and voted against the measure, but they did not "object strenuously or at length" against the proposal (Asher, 1991, p. B13).

At the 1992 Convention, the membership was presented with four agenda items sponsored by the Commission and the NCAA Council (on behalf of the Academic Requirements Committee), Proposals 14–17 were related to the initial eligibility changes proposed by the Commission the previous July. The two key items were No. 14 and No. 16. No. 14 stipulated the increase in core courses from 11 to 13, with the additional two courses to be taken in English, mathematics, or natural or physical science. No. 16 called for the establishment of the sliding scale index. In bringing No. 14 to the floor, Turner of Mississippi cited the educational organizations that supported the measure, including the Knight Commission, American College Testing, and the College Board. With little floor debate, No. 14 passed, 312 to 6, with one abstention (*Proceedings*, 1992).

Wake Forest president Thomas Hearn, a member of the Presidents Commission, brought No. 16 to the floor, noting the support of those same organizations. The proposed index, said Hearn, avoided "excessive reliance on a particular test score" (*Proceedings*, 1992, p. 233), a response to a consistent constituent claim that such standards disproportionately denied access to minorities. Hearn also characterized the public nature of the proposal:

> It is not an exaggeration to say that the nation is watching to see whether the Association will be part of the national effort to raise academic standards and to affirm above all the promises of the educational mission of the university as it pertains to student-athletes (*Proceedings*, 1992, p. 233).

Edward Fort of North Carolina A & T argued against the passage of No. 16: "It is my contention that insufficient time has expired relative to the availability of ultimate research results that could lead to a conclusion on an absolute basis; . . . it is not broken, so why are we attempting to fix it?" (*Proceedings*, 1992, p. 234). Said Georgetown President Rienzo:

> I am told we are looking for appropriate national standards, which I personally believe is impossible to attain in this complex heterogenous world of American higher education. . . . I am really concerned that we are proceeding based upon what is politically correct rather than what is academically sound (*Proceedings*, 1992, p. 238).

No. 16 was passed, 249 to 72, with five abstentions, with the provision that the new index would be in effect for incoming freshmen in the fall of 1995 (*Proceedings*, 1992).

Conclusion

The efforts of tethering intercollegiate athletics to institutional academic missions developed related consequences that continue to endanger the existence of initial academic eligibility. The criticisms that initial eligibility negatively and disproportionately impacts minorities has challenged these rules on the very basis on which they were founded. How can these rules meet the goal of tethering athletics with the mission of American higher education, these critics offered, if they serve to limit the access to higher education to those groups who have been historically and perniciously denied such access? Is this the goal of our higher education system, to limit access? While many continue to call for the repeal of eligibility rules based on standardized test scores, the strongest threats to these rules have come through legal challenges.

The most recent development in the ongoing struggle involves a lawsuit brought against the NCAA by four African American student-athletes who were denied Division I athletic eligibility due to existing legislation, and highlights the fact that this debate had become further enmeshed with the politics of racial discrimination. In March 1999, supporting the original criticisms leveled against No. 48 in the early 1980s, Judge Ronald L. Buckwalter decided in *Cureton et al. v. NCAA* that initial eligibility legislation does indeed have "an unjustified disparate impact against African-Americans," and identified statements in a 1998 NCAA memorandum that acknowledged that "African-American and low-income student-athletes have been disproportionately impacted by Proposition 16 standards. . . . For both (groups), the single largest reason for not meeting Proposition 16 standards was a failure to meet the minimum standardized test score" (*Cureton et al. v. NCAA,* p. 2). The court also stated that the NCAA "failed to articulate in any meaningful manner the decision making process behind the selection of the 820 cutoff score" (*Cureton et al. v. NCAA,* p. 22), (even though, as noted by Zingg [1983], the initial 700 SAT cutoff [raised to 820 by ETS when the scoring of the test was recentered] "represented the midpoint for [African Americans]" [p. 7]).

In response to the claim that the goals of current rules were to raise student-athlete graduation rates and close the gap between black and white student-athlete graduation rates, the court did not criticize the fact that the effort was made to raise graduation rates. "There appears to have been a perception that student-athletes were less academically prepared than the rest of the student body," said the court, adding, "Certainly, a public relations benefit would redound to the NCAA for having promulgated academic standards to combat these stories of abuse and exploitation. However, merely because a public relations benefit exists does not render the NCAA's adoption of (initial eligibility legislation) invalid" (*Cureton et al. v. NCAA*, p. 16). However, the court found that "not only is there no support for an educational institution to engage in (closing the gap between white and black student-athletes), but the proffered goal was unequivocally not the purpose behind the adoption of the initial eligibility rules" (*Cureton et al. v. NCAA,* p. 13). The court also stated that "the SAT has only been validated as a predictor of first-year GPA, and not college graduation" (*Cureton et al. v. NCAA,* p. 18), even though since 1986 and the advent of Proposal No. 48, graduation rates for all athletes, especially African Americans, have risen considerably (Suggs, 1999b). The court decision also highlighted three possible eligibility alternatives to current NCAA legislation put forth by the NCAA to eliminate the disproportionate impact. One model allows partial qualifiers full qualifier status, adjusting the sliding scale accordingly. A second extends the sliding scale ever further, while a third extends the scale still further and eliminates the minimum test score component entirely. The *Cureton* decision essentially verified the basic complaint that many critics (and some presidents) had leveled against current initial eligibility legislation: It disproportionately and inappropriately impacted African Americans. Also, Buckwalter's opinion supported the criticism that initial academic eligibility legislation was merely an attempt to wed athletics and academics to preserve the market value of the on-field sport product with sport-loving constituents.

Before the ruling in *Cureton,* although the absolute effectiveness of initial eligibility legislation in meeting stated goals had not been ascertained definitively, presidents who supported the current rules, understanding the importance of tethering to many constituents, had been unwilling to entertain a relaxation of the standards. Kenneth Shaw, Syracuse University chancellor and former chair of the Division I Board of Directors, said, "If we responded to every threatened lawsuit by changing practices, [the NCAA would] be a funny-looking place. We can't lower our expectations of students. What's important is we ought to respond to the facts and data, not criticism" (Carey & Milhoces, 1998, p. 2C). Shaw's response is ironic, as such criticism was a significant motivation for the implementation of the initial eligibility standards in the early 1980s in an effort to tether athletics and academics, and in that the findings in *Cureton* indicated that facts and data were not properly employed in setting current SAT cutoff standards.

Unless the courts continue to rule against initial eligibility in its current form, regardless of the positive perceptions of these standards, and unless the calls of racial discrimination ring more loudly, Shaw and his colleagues should continue to carry the standard of initial eligibility through the mechanisms of NCAA. While it seems that some form of initial eligibility will remain in the wake of *Cureton,* it is unclear at this writing as to what that form will be. On March 30, 1999, the 3rd Circuit Court of Appeals allowed the NCAA to keep pre-*Cureton* legislation in effect pending the outcome of its appeal, but the form of subsequent legislation depends not only on the courts, but also public opinion. If public opinion remains favorable, presidents will ensure that initial eligibility remains to whatever degree the courts allow. Following the *Cureton* decision, Penn State president Graham B. Spanier, current chair of the Division I Board of Directors, said:

> Almost everybody believes that academic integrity is extremely important and that some form of uniform standard for initial eligibility is necessary. Today, on average, student athletes graduate at a higher rate than the rest of the student body. That is a goal higher education should not abandon. (Suggs, 1999b, p. 149)

In addition, current NCAA executive director Cedric Dempsey indicated that he believed no president would look favorably on admitting any prospective student-athletes with extremely low standardized test scores (Suggs, 1999b). These responses would seem to indicate that many presidents still feel that there is power in their stand on initial eligibility. However, some post-*Cureton* proposals eradicate any reliance on testing and GPAs and instead focus on penalizing programs where student-athletes leave school due to academic difficulties. A plan coauthored by Vanderbilt chancellor Joe B. Wyatt would prohibit schools from reassigning the grant-in-aid of a student-athlete who permanently loses eligibility until that student-athlete's class graduates. Said Wyatt: "If you value educating student-athletes, you should favor this plan. Presidents, chancellors and faculty will have a hard time arguing against it" ("NCAA reform," 1999, p. 32). The value of such a plan is that it makes no use of those elements identified by the court in *Cureton* as discriminatory against African Americans.

While the immediate future of initial eligibility legislation is unclear, what is clear is that presidential efforts significantly influenced the NCAA's adoption of related rules. The early calls for reform by those like Graham, though well intentioned, failed in large part due to a lack of consensus and lack of coordinated action on the part of presidents. The fear of engendering competitive on-field disadvantages, coupled with the resulting backlash from sport-loving constituents, and the ramifications of the eventual shifting constituent opinions toward minority access to higher education (as supported by Federal legislation), drove many in the NCAA to dismantle first the Sanity Code, then the 1.600 Rule. However, as the national interest in intercollegiate athletics grew, especially in football and men's basketball, so too did the eventual constituent recognition that some student-athletes were not able, willing, required or expected to actually be students. This realization challenged the ideal notions of intercollegiate athletics and of the student-athlete.

As a result, as with the establishment of No. 48, presidential supporters of tethering found an issue around which they could build consensus amongst themselves to act in consort to attempt to establish athletic and academic congruence.

The question of whether intercollegiate athletics can ever be fully in harmony with academics in the constituent-based systems of American higher education has been debated since intercollegiate athletic programs were formally incorporated by institutions. From the earliest formations of the NCAA, many presidents have taken an active role through the association in seeking, using Helman's terms, to tether intercollegiate athletics to the academic mission of their institutions. Interested presidents have attempted to affect change through avenues other than the NCAA such as ACE, but as the NCAA solidified its hegemony in intercollegiate athletics governance, presidents pursued changes through a greater and more formalized involvement in the organization.

As the NCAA's decision-making process has been altered by presidential involvement, much in the way of leadership concerning initial eligibility legislation has been ceded to interested presidents. As this debate continues, presidents, who are seen by many who participate in this debate to be the individuals who must provide an institution and its constituents with an articulated vision for its future, are now more likely to be expected to address these questions rather than athletic directors, coaches, or NCAA staff. However, while these presidential efforts in establishing initial eligibility legislation are seen by most as supportive of congruence, it has been at the expense of constant criticisms that African Americans bear the brunt of the restrictions that stem from these rules. Despite these criticisms and the potential reversal dealt by the *Cureton* decision (if upheld), involvement in proposing and determining academic legislation has allowed presidents to assume a generally popular stance with most constituent groups.

In reflecting on presidential control over intercollegiate athletic policy at the end of the century, sport sociologist D. Stanley Eitzen summarized the expectations of many critics, calling for "a significant reform effort led by college presidents that would clean up college sports programs so that they are consonant with the educational objectives of the institutions they represent.... (But) the presidents of the universities involved in big-time college sport are too weak or too meek or too unwilling to change" (1999, pp. 124, 126). However, as Bok pointed out, just what those objectives are is open for interpretation, and the interpretation process is influenced by a great many constituent groups. In the case of academic legislation, many presidents have taken an active role in addressing the congruence issues as framed by Eitzen, but it is also with the realization that for intercollegiate athletics to survive and to meet market demands, athletic and academic congruence at some level must be acknowledged. If interested presidents preserve their current approaches concerning initial eligibility and craft through the NCAA a version of initial eligibility legislation that withstands the legal scrutiny of the courts, then collectively they may be able to prove Eitzen's claims of weakness wrong, and the tethering of academics and athletics will stand. However, as noted above, the NCAA's restructuring could make direct presidential impact more difficult to exert, shifting the balance of power away from presidents even on issues pertaining to academics. In addition, related controversies continue to emerge, such as the allegations of a tutor who claimed to have written—with the knowledge of athletic department academic counselors—400 papers for men's basketball players at the University of Minnesota, along with claims that former head coach Clem Haskins intimidated faculty into giving his players special consideration (Wertheim & Yaeger, 1999).

It is situations like these that threaten the popularity of the game product with constituents. As these situations arise, interested presidents will continue to be pressed to deal with the fallout, to craft responses that are in line with the mission of their institution, and to assuage the criticisms from within and outside academe. These efforts to tether academics and athletics, while deemed vital by many, less so by others, are still a crucial component in the viability of meeting NCAA standards and securing the popularity of the game product by maintaining ideal notions of intercollegiate athletics and the student-athlete. For it is clear that the constituent system demands that schools provide not only winning teams, but teams that win within constituents' ideal framework of intercollegiate athletics. These demands will continue to challenge presidential leadership capacities and their abilities to provide appropriate institutional vision and to deal with the calls for congruence within the constituent-based American higher education system.

References

1.600 rule modified but not abolished by 65th convention. (1971, February 1). *NCAA News,* p. 3.

'90 convention will get legislation to postpone initial-eligibility changes. (1989, January 25). *NCAA News,* pp. 1–2.

125 CEOs are expected. (1983, January 5). *NCAA News, 20*(1), p. 7.

1964–1965 yearbook of the National Collegiate Athletic Association. (1965). Kansas City: National Collegiate Athletic Association.

1989 NCAA convention proceedings. (1989). Mission, KS: National Collegiate Athletic Association.

1990 NCAA convention proceedings. (1990). Overland Park, KS: National Collegiate Athletic Association.

1991 NCAA convention proceedings. (1991). Overland Park, KS: National Collegiate Athletic Association.

1992 NCAA convention proceedings. (1992). Overland Park, KS: National Collegiate Athletic Association.

1999–2000 NCAA Division I operating manual. (1999). Indianapolis, IN: National Collegiate Athletic Association.

A history of academic legislation. (1980, December 15). *NCAA News,* pp. 2–3.

Ames, L. (1989, February 28). Why use SAT? No alternative. *Boston Globe,* pp. 63, 68.

Are academic qualifications strong enough? (1980, April 15). *NCAA News,* pp. 3, 6.

Asher, M. (1991, June 30). NCAA Presidents' academic issues: 'Education, politics or discrimination?' *Washington Post,* pp. B1, B13.

Bailey, W. S. (1986, April 30). Bailey reviews decade preceding Bylaw 5-1–(j) requirement. *NCAA News,* p. 3.

Bannon, J. (1989, January 17). Thompson is likely to continue boycott. *USA Today,* p. 1C.

Barnhart, T. (1989, June 15). Conference executives praise SEC move on nonqualifiers. *NCAA News,* p. 5.

Betts, J. R. (1974). *America's sporting heritage: 1850–1950.* Reading, MA: Addison-Wesley.

Bok, D. (1983, December 14). Presidents need power within the NCAA to preserve academic standards and institutional integrity. *Chronicle of Higher Education, 27*(16), pp. 33–34.

Bok, D. (1985a). Intercollegiate athletics. In J. B. Bennett & J. W. Peltason (Eds.), *Contemporary issues in higher education* (pp. 123–146). New York: Macmillan.

Bok, D. (1985b). Presidents need power within the NCAA to preserve academic standards and institutional integrity. In D. Chu, J. O. Savage, & B. J. Becker (Eds.), *Sport and higher education* (pp. 207–210). Champaign, IL: Human Kinetics Publishers.

Brubacher, J. S., & Rudy, W. (1976). *Higher education in transition: A history of American colleges and universities 1636–1976.* New York: Harper and Row.

But for Thompson, 42 would have survived, Schultz says. (1989, February 1). *NCAA News,* p. 16.

Byers, W., with Hammer, C. (1995). *Unsportsmanlike conduct: Exploiting college athletes.* Ann Arbor, MI: University of Michigan Press.

Cal suit dismissed, action void. (1972, May 15). *NCAA News,* p. 2.

Carey, J., & Milhoces, G. (1998, June 30). Organization in transition. *USA Today,* pp. 1C–2C.

Convention acts on key athletic issues. (1983, January 19). *NCAA News,* pp. 1, 12.

Cureton, et al. v. NCAA, 37 F. Supp. 2D 687; 1999 U.S. Dist. LEXIS 2359.

Davidson, D. (1991, June 30). Mixed reviews for tougher Prop 48. Atlanta Constitution, p. E15.

Douchant, M. (1989, January 23). Thompson protests stiffer NCAA requirements. *The Sporting News,* p. 30.

DiBiaggio, J. A. (1995, Summer). Attitudes about winning. *The Long Term View, 3*(2), pp. 8–13.

Edwards, H. (1983, August). Educating black athletes. *Atlantic Monthly, 225*(2), pp. 31–38.

Eitzen, D. S. (1999). *Fair and foul: Beyond the myths and paradoxes of sport.* Lanham, MD: Rowman & Littlefield.

Falla, J. (1981). *NCAA: Voice of college sports.* Shawnee Mission, KS: National Collegiate Athletic Association.

Farrell, C. S. (1984, September 5). Sport rule will bar able blacks, NCAA study says. *Chronicle of Higher Education,* pp. 1, 31.

Farrell, C. S. (1985a, April 10). New academic rules for freshman athletes assailed as 'artificial barrier.' *Chronicle of Higher Education,* p. 33.

Farrell, C. S. (1985b, November 15). NCAA Council and presidents' panel agree to back 3-year phase-in for academic rules. *Chronicle of Higher Education,* pp. 32–33.

Fiske, E. B. (1983, January 14). Athletes' test scores. *New York Times,* p. 71.

Fleisher, A. A., Goff, B. L., & Tollison, R. D. (1992). *The National Collegiate Athletic Association: A study in cartel behavior.* Chicago: The University of Chicago Press.

Funk, G. D. (1991). *Major violation: The unbalanced priorities in athletics and academics.* Champaign, IL: Leisure Press.

Governance approval highlights convention. (1981, January 31). NCAA News, pp. 1, 6.

Grade-point impact. (1991, July 3). *USA Today,* p. 9C.

Gross, J. (1982, March 24). Reformers try to halt academic abuses. *New York Times,* p. B9.

Hanford, G. H. (1985). Proposition 48. In D. Chu, J. O. Savage, & B. J. Becker (Eds.), *Sport and higher education* (pp. 367–372). Champaign, IL: Human Kinetics Publishers.

Helman, J. W. (1989). *A history of American intercollegiate athletic eligibility: Educational compromises to competitive interests.* Unpublished doctoral dissertation, Pennsylvania State University.

Jenkins, D. (1966, March 21). It's one point six, pick up sticks. *Sports Illustrated,* pp. 30–31.

Knight Foundation Commission on Intercollegiate Athletics. (1991). *Keeping the faith with the student-athlete: A new model for intercollegiate athletics.* Charlotte, NC: Knight Foundation.

Lawrence, P. L. (1987). *Unsportsmanlike conduct: The National Collegiate Athletic Association and the business of college football.* New York: Praeger.

Lawson, H. A., & Ingham, A. G. (1980, Winter). Conflicting ideologies concerning the university and intercollegiate athletics: Harper and Hutchins at Chicago, 1892–1940. *Journal of Sport History, 7*(3), 37–67.

Lilley, T. J. (1987, September 28). Effects of Bylaw 5-1-(j) won't be known for years. *NCAA News,* pp. 1–2.

Minimum academic standards in wind. (1964, July-August). *NCAA News,* pp. 1, 4.

Missanelli, M. G. (1990, January 9). NCAA reaches compromise on Proposal 42. *Philadelphia Inquirer,* pp. 1F, 4F.

Mission statement (2000). American Council on Education. [On-line]. Available at: http://www.ACENET.edu/About/mission.html.

Modifications in academic standards could be announced later in month. (1985, July 17). *NCAA News,* p. 1.

Moore, D. L. (1989, January 17). 2 black schools say votes belie positions. *USA Today,* p. 6C.

Morris, W. (1992). *The courting of Marcus Dupree* (Rev. ed.). Jackson, MS: University of Mississippi Press.

Mott, R. D. (1995, January 4). The 100-year debate. *NCAA News supplement: 89th convention issue,* pp. 1, 10, 11.

NCAA debates academic-standards plan. (1986, January 13). *Boston Globe,* p. 34.

NCAA reform: The Vandy plan (1999, June 21). *Sports Illustrated,* pp. 30–32.

NCAA to seek delay on entrance test rule. (1984, October 19). *New York Times,* p. B16.

New academic rules would hit blacks hardest. (1984, September 10). *NCAA News,* pp. 1, 16.

New "triple option" may be submitted. (1978, June 1). *NCAA News, 15*(10), p. 4.

Oberlander, S. (1989, January 25). New NCAA rule on aid stirs emotional protest; repeal seen possible. *Chronicle of Higher Education,* pp. A1, A36.

Opinions out loud. (1977, May 15). *NCAA News, 14*(7), p. 2.

Opinions shifting on Proposition 42. (1989, February 17). *New York Times,* p. A32.

Proceedings of the 66th annual convention of the National Collegiate Athletic Association (1972). Kansas City: National Collegiate Athletic Association.

Proceedings of the 77th annual convention of the National Collegiate Athletic Association (1983). Mission, KS: National Collegiate Athletic Association.

Proceedings of the 80th annual convention of the National Collegiate Athletic Association (1986). Mission, KS: National Collegiate Athletic Association.

Purity code rejected. (1948, August 10). *New York Times,* p. 28.

Results of survey on bylaw 5-1-(j) released. (1988, April 20). *NCAA News,* pp. 1–2.

Rhoden, W. C. (1989a, January 13). N.C.A.A. is criticized for scholarship change. *New York Times,* p. A25.

Rhoden, W. C. (1989b, January 21). Thompson returns as N.C.A.A. acts to delay rule. *New York Times,* p. A23.

Robbins, D. (1990, January 7). NCAA will look at Proposition 42 again. *Los Angeles Times,* p. 13.

Rudolph, F. (1990). *The American college and university: A history.* (Rev. ed). Athens, GA: University of Georgia Press.

Sack, A. L. (1986, November 9). Looking behind the Proposition 48 image. *New York Times,* p. 10.

Sack, A. L., & Staurowsky, E. J. (1998). *College athletes for hire: The evolution and legacy of the NCAA's amateur myth.* Westport, CT: Praeger.

Schuster, R. (1989, January 18). Thompson won't budge; NCAA could. *USA Today,* p. 1C.

SEC planning legislation to ban grants to partial qualifiers. (1988, August 31). *NCAA News,* p. 3.

Smith, R. A. (1988). *Sports and freedom: The rise of big-time college athletics.* New York: Oxford University Press.

Special committee's report on academic criteria. (1985, September 9). *NCAA News, 22*(31), p. 12.

Sperber, M. (1998). *Onward to victory: The crises that shaped college sports.* New York: Henry Holt.

Suggs, W. (1998, December 18). NCAA's new structure finds fans and foes among sports officials. *Chronicle of Higher Education,* pp. A37–A38.

Suggs, W. (1999a, January 22). Apathy prevails at NCAA Convention as fewer officials choose to attend. *Chronicle of Higher Education,* pp. A35–A36.

Suggs, W. (1999b, April 9). Fight over NCAA standards reflects long-standing dilemma. *Chronicle of Higher Education,* pp. A48–A49.

Telander, R. (1989). *The hundred yard lie.* New York: Simon and Schuster.

Testing service challenges NCAA's No. 42. (1989, January 18). *NCAA News,* pp. 1–2.

Thelin, J. R. (1996). *Games colleges play: Scandal and reform in intercollegiate athletics.* Baltimore: The Johns Hopkins University Press.

Thompson to walk in protest of NCAA. (1989, January 14). *Boston Herald,* p. 69.

Two changes sought for Proposal No. 48. (1985, August 28). *NCAA News,* pp. 1, 12.

Vance, N.S. (1984, January 18). Presidents' commission set up by NCAA, but ACE plan loses. *Chronicle of Higher Education,* pp. 1, 34.

Vote confused. (1989, January 18). *USA Today,* p. 7C.

Weaver, W., Jr. (1989, January 14). Thompson to leave bench in protest on scholarships. *New York Times,* p. 45.

Wertheim, L. J., & Yaeger, D. (1999, June 14). The passing game. *Sports Illustrated,* pp. 92–102.

White, Jr., G. S. (1980, October 15). Preferred admissions reported at U.S.C. *New York Times,* pp., B7–B8.

Wieberg, S. (1986, January 13). Academic bill approval expected. *USA Today,* p. 6C.

Wieberg, S. (1989a, January 12). Proposal 48 toughened. *USA Today,* p. 1C.

Wieberg, S. (1989b, October 19). Schultz: Prop 42 to change. *USA Today,* p. 5C.

Wieberg, S. (1991, May 8). Junior colleges, prep schools become sanctuary. *USA Today,* p. 3C.

Yaeger, D. (1991). *Undue process: The NCAA's injustice for all.* Champaign, IL: Sagamore Publishing.

Zingg, P. J. (1983, Summer). No simple solution: Proposition 48 and the possibilities of reform. *Educational Record, 64*(3), pp. 6–12.

Michael R. Mills
Adrienne E. Hyle

No Rookies on Rookies

Compliance and Opportunism in Policy Implementation

It is common for writers and researchers to note that scrutiny of American higher education institutions has been steadily increasing in recent years. The general public and its elected and bureaucratic representatives are increasingly willing to be pointed and critical of colleges and universities. These external stakeholders are no longer willing to defer to the faculty and institutional administrators as they enforce policies and procedures that some find questionable. One example of this phenomena is the concern of the general public about the quality of undergraduate instruction and the related calls to replace graduate teaching assistants with senior faculty in freshmen-level courses. While the procedures for assigning instructors to first-year classes vary from institution to institution and department to department, in many cases colleges and universities have tried to respond to the public's demands. There have been greater efforts to provide training for graduate teaching assistants and particularly international teaching assistants. There have also been changes in who shoulders the teaching responsibility for freshmen-level courses.

This study looks at one such effort at Oklahoma State University (OSU). The university is one of two flagship universities in the state system of public higher education and has a Carnegie designation of Research University Extensive. It has proudly retained its traditional land-grant mission and image and is located in a town about 60 miles from the state's two major metropolitan centers, Oklahoma City and Tulsa. The

Michael R. Mills is assistant professor, Department of Leadership Development, University of South Florida, and Adrienne E. Hyle is professor at Oklahoma State University.

The Journal of Higher Education, Vol. 72, No. 4 (July/August 2001)
Copyright © 2001 by The Ohio State University

main campus, which is the focus of this study, enrolls roughly 20,000 students. At the time of our study, the faculty of almost 1,000 was organized into 58 academic departments arrayed among seven colleges.

In 1994 the new president of the university, James Halligan, made increasing enrollment one goal of his administration, and he immediately began to refocus on student recruitment and retention strategies. In response to students' complaints and public criticism of the university's freshmen-level instruction and after limited and brief discussion within the OSU community, the president announced that inexperienced and international graduate students would no longer be allowed to instruct freshmen-level courses.

The Policy

The policy was labeled "No-Rookies-on-Rookies" (NROR). From its inception, President Halligan had indicated his hope that the NROR program would improve classroom experiences for freshmen. In a memo to department heads and deans, he indicated that they should "develop and implement enhanced procedures to insure experienced teachers who possess effective communication skills are present in our freshmen classes" (Halligan, memo, 1995, p. 2). Although "the policy's fine print remains vague" ("Proposed Policy," 1995, p. 4), reports in the university newspaper based on interviews with the president and provost suggested that collateral goals of the program were to enhance freshmen students' comfort and satisfaction with their experiences at OSU, to increase freshmen retention, and to improve graduate students' preparation for teaching.

The policy was a restrictive mandate to colleges and academic departments, but it was not prescriptive. It told them what *not* to do, but the policy did not set specific solutions or strategies for implementation. Thus, colleges and academic departments were allowed to formulate their own strategies to compensate for no longer assigning freshmen-level teaching responsibilities to rookie and international graduate students. The policy intent was that rookie graduate students undergo year-long training programs designed to introduce them to the realities of university-level instruction. OSU's president committed additional funding for the continued employment of rookie and international graduate students in upper-level teaching positions or other duties. Departments devised various solutions for the staffing change based on their differing interpretations of the policy, the nature of the problems the policy created for them, and the alternative uses of graduate students and faculty that fit their circumstances.

Despite this flexibility, the new policy meant changing long-established approaches to staffing freshmen-level courses. Throughout the nation, graduate students are used as instructors in lower-level undergraduate recitation and laboratory classes. They cover entry-level classes, thereby freeing up faculty to teach coursework for majors and graduate students in the discipline and to focus on scholarly and service activities. Graduate teaching assistants (GTAs) cost substantially less than full-time faculty. In addition to being cost effective, the use of graduate teaching assistants in lower-level undergraduate coursework can be justified from a professional socialization perspective. The experiences gained by graduate teaching assistants in the classroom are necessary for their development as future faculty, and many hiring institutions prefer candidates with a variety of teaching experiences prior to graduation ("Colleges Expand," 1992; Connelly, 1982). In sum, implementation of the NROR policy had substantial ramifications for departments beyond its impact in the freshmen-level classroom.

Purpose and Research Questions

The purpose of this project was to examine the implementation of the NROR policy. Specifically, we sought answers to the following questions:

1. What did departments change in their implementation of the NROR policy?
2. What factors influenced departmental responses to the NROR policy? What role did public concerns play? What role did unit contexts play?
3. What advice can be generated for those designing institutional responses to public concerns?

Related Literature

The research on policy implementation confirms the obvious: policy does not get translated directly into organizational actions. Implementation requires additional rounds of interpretation and negotiations at the implementer's level. At this level, issues and concerns may be very different, and the beliefs and values upon which decisions are made may not necessarily coincide with those of policy makers (Baier, March, & Saetren, 1986; March, 1994; Mills, 1998). Thus, assumptions and concepts applied in the literature on organizational decision making and policy implementation provide frameworks for viewing and analyzing responses to policies.

Matching themes in the literatures of both fields center on a turn to interpretive analyses based on the assumption that people encounter circumstances that can be understood in multiple ways. People must interpret, or make sense, of those situations in order to determine how to act (Weick, 1995). They must also compare their interpretations with other people to see if they provide a reasonable foundation for mutual understanding and coordinated action (Fischer & Forester, 1987, 1993; Yanow, 1996). In other words, interpretations are applied and developed socially, and they depend heavily on the contexts in which the social interactions occur. Organizations, as social collections in which people must interpret in order to facilitate action, embody continuous exercises in sensemaking and often have contexts for sensemaking nested within contexts (Linstead, Small, & Jeffcutt, 1996; Pondy, Frost, Morgan, & Dandridge, 1983; Weick, 1995).

Interpretive processes are often particularly apparent when organizations undertake implementation of a new policy (Fischer & Forester, 1987, 1993; Yanow, 1996). In fact, the processes related to implementation are the result of negotiated interpretations of the letter and spirit of policy (Bowe, Ball, & Gold, 1992).

> An interpretative approach to policy analysis, then, is one that focuses on the meanings of policies, on the values, feelings, and/or beliefs which they express, and on the processes by which those meanings are communicated to and "read" by various audiences. (Yanow, 1996, pp. 8–9).

In fact, as Ball makes clear,

> Solutions to the problems posed by policy texts will be localised and should be expected to display ad hocery and messiness. Responses indeed must be "creative". . . . Policies do not normally tell you what to do, they create circumstances in which the range of options available in deciding what to do are narrowed or changed or particular goals or outcomes are set. A response must still be put together, constructed in context, off-set against other expectations. All of this involves creative social action not robotic reactivity. Thus, the enactment of texts relies on things like commitment, understanding, capability, resources, practical limitations, cooperation and (importantly) intertextual compatibility. (Ball, 1997, p. 10)

In some cases, this recreation results in an "interpretation of texts [that] is proactive, critical and self-assured, . . . in others, reactive, passive and unquestioning" (Bowe et al., 1992, pp. 119–120).

This aspect of policy implementation, when considered as an avenue that allows divergence from policymakers' intent, has been viewed in two ways. Some lines of research aim to discover ways to limit policy reinterpretation (Mazmanian & Sabatier, 1983). However, other researchers consider implementation reinterpretation as part of the policy process (Palumbo & Calista, 1990) that allows policymaking to be symbolic action demonstrating values while ensuring that implementation serves the interests of the various constituencies, thus serving to maintain a balance of power (Ferman, 1990).

In addition to issues of policy reinterpretation, the processes of organizational decision making in response to policy pronouncements further impact policy implementation. Cohen and March (1974), March and Olsen (1979) and March (1994) provide a framework for analyzing decision making that is consistent with the interpretive assumptions discussed above.

They remind us that, in situations with multiple, uncertain preferences and unclear technology, success is often difficult to determine, and learning, therefore, has uncertain foundations. Combine this with decision processes in which participants' attention is limited and involvement is fluid, and it makes ambiguous conditions for decision making. These analysts characterize organizations with these features as organized anarchies that exemplify garbage can decision-making processes. In such circumstances, problems, solutions, and decision makers float through the organization, and decision making occurs when they come together in particular choice opportunities. Options abound, and outcomes are often unintended and various. "The 'decisions' of the system are a consequence produced by the system but intended by no one and decisively controlled by no one" (Cohen & March, 1974, p. 34). Instead, given the ambiguity organizational actors face in choice opportunities such as policy implementation, instances of decision making become "vehicles for constructing meaningful interpretation" (March, 1994. p. 199) that promote sensemaking, explain some of the ambiguity, and promote choices. In organized anarchies, we learn about who we are and what we value by seeing what we decide to do.

The typical tactics of administrative action in an organized anarchy are based on the assumption that the responsibility of leaders is to prompt the organization to a changing and more complex view of itself. Because decisions demand interpretation, they become opportunities to define the organization and its connections to its world. From this foundation, March and Olsen (1979) develop eight suggestions for those who seek to influence the course of decisions in universities and colleges: (1) devote time and energy, (2) persist, (3) get things done any way you can, (4) facilitate opposition participation, (5) engage in lots of projects, (6) provide garbage cans, (7) manage unobtrusively, and (8) interpret history.

Procedures and Method

We began by sending a letter to each department head asking if the NROR policy affected one or more of the unit's classes. Of those who responded, only 10 department heads reported a direct impact from the NROR policy. A check of the University catalog indicated that only one of the nonrespondents headed a department that taught a freshmen-level course. We contacted the 11 heads and arranged to interview them or someone they designated as the person responsible for their departments' implementation of the policy. In addition, we interviewed college-level administrators from the two colleges whose unit heads indicated some impact from the policy.

We talked to these administrators during the spring of 1997, the second year of the NROR policy implementation. We conducted semistructured interviews that allowed participants to discuss the issues they thought were most important and permitted us to clarify points through follow-up questions and probes (Kvale, 1996). With each participant in the study, we asked about college and department instructional resources, GTA training, international GA placements, and freshmen-level course design both before and after the implementation of NROR. We also asked how each department went about deciding what to do to meet the policy's expectations—how decisions were made, who was involved in those decisions and how they participated, what the most salient considerations were, and what type of reporting about those decisions and subsequent actions were required to higher levels of the college and university. Further, we asked how the implementation had changed from the first year to the second.

The interviews were audio tape-recorded and transcribed verbatim. The transcripts were coded by assigning a phrase summarizing the content to distinct pieces of the text. The codes served as convenient markers and facilitated analysis through comparative clustering of related textual units that helped us develop a fuller understanding of departments' approaches to implementing the policy.

Not all of the heads interviewed actually had a direct role in the implementation of the policy. For example, one head interviewed with us in order to express his pique at not being able to share in the resources made available to compensate for implementation of NROR. His department had never used rookies in freshmen-level classes, and, from his viewpoint, other departments were getting compensatory funding to do what his department had done correctly all along and without the extra money. Leaving aside this and like cases, we were left with seven departmental-level interviews (five heads, one assistant head, and one graduate coordinator at the time of implementation and three college-level interviews (involving a dean and two associate deans) as the bases for our analyses. To protect confidentiality of the study participants, we changed all department and discipline references to more generic descriptors (for example, “Introduction to Alchemy” became “the freshmen-level course”).

Implementation Changes

Implementation of the NROR policy occurred in one college and seven departments of a second college, referred to here as College 1 and Departments A–G of College 2. Table 1 summarizes the changes the colleges and departments made as part of the NROR policy.

College 1 is a professional college that relies on the arts and sciences college to instruct its freshmen in the basic subjects prerequisite to studies in its majors. The leadership of College 1 had grown increasingly concerned that its faculty and staff were not more involved with its new students and that the college suffered much of its student attrition before students began course work in the college.

> I think we needed to have that contact with the freshmen year. We’ve always regretted the fact that we turn our students over to Arts and Sciences. We’ve always had this concept that you’ve got to build this foundation, so we have limited exposure to the students. We lose half our students in those first two years, and, so, what can we do about that?

The administration of the college was already in the process of trying to address this concern when the president instituted the NROR policy. The policy became the spur that caused the administration of the college to accelerate its planning and decision making. They initiated a freshmen seminar in which faculty and staff met regularly with small groups of students (around 20) to discuss the professional field as well as topics generic to students’ experiences in the university. As one participant described the seminar design,

> Each section could have its own focus. We also, then, offered a series of seminars that were on topics of general interest to all the sections. Specifically, we have two seminars on what is [this profession]—all the different fields, one on study and test-taking skills, one on time management, an assessment of computer skills, and a session on preparing a resume.

The administration of the college believed that the key to the new offering was not only what was studied, but also that its new students had an additional resource to aid their adjustment to the university. As we were told, “It’s the concept of, ‘I want to talk to each one of you today, and each of you should have a chance to talk in the session’ versus ‘come in, sit down and listen.’ They want students to talk and participate and get involved.”

In College 2, the task of coordinating the implementation of NROR

TABLE 1

1000-Level Instruction Before and After NROR Implementation

Unit	Before Policy	After Policy
College 1		
1000-level instructors	None	Full-time faculty, administrators and staff
Classes and size	None	Freshmen Experience Seminar @ 20 per section
New GTA training*		N/A
College 2—Department A		
1000-level instructors	GTAs	Visiting faculty members
Classes and size	Recitations @ 35	Multiple recitations @ 100
New GTA training		Faculty mentors, seminar on teaching
College 2—Department B		
1000-level instructors	GTAs	Full-time faculty
Classes and size	Recitations @ 60–80	Recitations @ 200
New GTA training		N/A
College 2—Department C		
1000-level instructors	GTAs	Full-time faculty
Classes and size	Recitations @ 40	Recitations @ 120 (1st year) 60–70 (2nd year)
New GTA training		N/A
College 2—Department D		
1000-level instructors	GTAs	Experienced TAs and part-time instructors
Classes and size	Recitations @ 25–30	No change
New GTA training		Mentoring program, rookies assigned to assist students in an individual help center
College 2—Department E		
1000-level instructors	GTAs	Part-time instructors
Classes and size	Labs	No change
New GTA training		One week orientation, sessions from COE program
College 2—Department F		
1000-level instructors	GTAs	Part-time instructors
Classes and size	Labs	No change
New GTA training		Faculty supervisor
College 2—Department G		
1000-level instructors	GTAs	Part-time instructors
Classes and size	Recitations @ 30	No change
New GTA training		Faculty group advisor

*New GTA training refers only to the time after policy implementation.

fell to the dean's office staff, who took on the responsibility for specifying what the policy meant and communicating that to the academic departments. Administrators in the dean's office approached this task in a direct manner. As we were told, NROR "has been imposed from the top down. There was no discussion once it got to our level and below. . . . We have some people who don't like it, who complain about it to some faculty members. I just tell them, 'It's not really an item for negotiation, this is the way it's got to be.'" However, various aspects of the policy needed clarification, which members of the provost's office provided.

> In the beginning, I had to make some phone calls back to the Provost's Office to get some decisions. One of them that came up was, what exactly is "experience?" They told me teaching experience can be student teaching, when you were an undergraduate. It can be having taught in a high school. . . . It doesn't have to be college teaching, it could be teaching at any level. . . . We also had to define what "international" was initially, and international was defined as someone who's from a country in which the native language is not English.

With this guidance, the dean's office staff developed the following statement to guide implementation of the policy:

> No 1000-level (freshmen) course will be taught by a "rookie" (someone with no teaching experience), no international teaching assistant will be teaching 1000-level courses and no 1000-level laboratory will be assisted by an international teaching assistant. "International" is defined as those students coming from countries whose official language is not English. (Meeting distribution from Dean's Office, n.d., original is in all capital letters)

One final component of NROR was that any graduate student who could not get an assistantship due to the policy would receive an equivalent award funded from a centrally held pool of funds, that is, not at the department's or college's expense. There was every willingness to fund as many of these nonteaching graduate students as it took to fairly implement NROR. Departments were allowed to rank their graduate students and then seek to fill whatever GTA slots they had available. Students on the list who could not fill any of the available positions due to NROR were bought out. As someone in the dean's office explained:

> We will give you money to move that [ineligible] person out, and you then buy another TA who is eligible to replace that person. Now the interesting question that came up was: Does that mean the next person I hire has to be somebody who qualifies? The answer to that is, no. The next person should be the next highest person in terms of qualifications or priority. We have done this so that they go down the list, and we are constantly buying out the next person on the list. There's no intent here to discriminate in terms of hiring. We've always been very emphatic about that. The next person you hire, doesn't necessarily have to be qualified to replace that person. We want you to hire the next person.

Despite the straightforward definitions of the key elements of the policy and the financial support that diminished some of the possible negative consequences of implementation, there was significant variety in implementation. What the departments of College 2 did in their implementation of the policy differed for several reasons. First, the policy as it was defined contained some distinctions that varied in applicability across the departments. For instance, rookie native English speakers were allowed to teach laboratory and discussion sessions as long as they were not the primary person responsible for teaching the course, but international students, even with teaching experience, were not allowed to teach in any capacity at the freshmen level. This meant that the departments with large numbers of international graduate students had more people permanently disqualified from freshmen teaching than other departments.

In addition, some disjunction in the policy was created because freshmen classes were defined as classes at the 1000-level. Several departments had courses intended for freshmen but with higher numbers. Thus, as one person noted,

> There are some departments in [College 2] that do not have a 1000-level class. That is really kind of, in my opinion, a fluke. It's absolutely a fluke. Introduction to [one subject] should be no different than Introduction to [another subject] or Introduction to [another], but those have a 2000 number. [Another department], I think, is the same way. There are a couple of departments who have no 1000-level classes. So they, in that sense, are out of it. They still have rookies. Because those departments' classes are at a 2000 level, you can have a brand new TA who's never taught.

In effect, some classes that seemingly fit the spirit of the NROR prohibition nevertheless slid out from under its application.

The second aspect of the variation in implementation concerned the differential application of the policy by the college. Unlike the five other implementing departments in the college, two departments (B & C, Table 1) were simply told that they were to replace their teaching assistants with full-time faculty members rather than nonrookie GTAs. Training graduate students and working them into responsibilities in the freshmen level courses was not an option for these two departments. As one department head told us, "The dean's office acted unilaterally in telling the department to have full-time faculty take over all the teaching in [the freshmen-level course]. They made the policy that there would be no graduate students in the classroom, and faculty should be assigned there." In fact, a dean's office representative indicated to us that one of these departments was not under the NROR policy, and instead viewed the changes there as a separate matter. The department head, however, thought that the changes they made were part of the NROR initiative, and the requirements for that department matched those placed on another department that all parties considered within the scope of NROR. One of the heads implied that the faculty-for-GTAs switch was the preferred implementation strategy of the dean's office, but that the dean's office could not apply it universally.

> The dean's office took a simple approach. However, they were not able to apply that simple an approach with [other] departments because they are too large and teach too many freshmen students to be able to fully staff the courses with full-time faculty. Thus, they were allowed to continue to use graduate students in their 1000-level courses.

With the changes required by the dean's office, both departments had to increase class size at the freshmen level significantly to comply with the limits on implementation they faced.

Of the other College 2 departments, one, Department A, used the NROR policy to accommodate extensive changes in its freshmen classes, while the other four, D through G, made few changes. All five of these departments also employed some form of training, including workshops, seminars, mentorships, and progressive, monitored experiences to enhance the teaching abilities of their rookie graduate students.

In Department A, the full-time faculty compensated for some of the lost GTA instruction by increasing their offerings of alternatives to the standard, freshmen-level course. Furthermore, the money freed up by the NROR funding for graduate students provided partial funding for hiring visiting faculty members who were assigned the typical freshmen class, albeit in larger sections than before the new policy.

The remaining four departments kept their freshmen-level courses operating in much the same way as before the NROR policy. They put some form of teacher training in place for new graduate students and relied on experienced GTAs and part-time instructors to teach the affected courses. One of these was a department with a large service course, and the head felt there were few viable options to maintaining the course as close as possible to the manner it had operated in the past.

> Unless we hire, you know, 30 new staff members, or 30 post docs who have experience, and don't worry about our graduate students—just let them be graduate students, and of course offer all kinds of fellowships and scholarships—basically, we at the university are stuck with putting people in the classroom who don't know an awful lot about teaching.

Another two of these departments traditionally used GTAs only in laboratory sections, and were affected primarily by the ban on international GTAs in that setting. They chose to use a fuller definition of rookies and not to use inexperienced native English speakers in their laboratories as well.

After taking an active role in interpreting and differentially applying the policy, the dean's office staff did little to govern or monitor the actions taken in the name of NROR. As a member of the office put it, "We distribute the funds and we try to answer questions—'Yes, this person can; no, this person can't.' " But, he added, "We don't dive into a department and look, but we tell them what they can and cannot do. We don't check up on them." With respect to training, he added,

> that is up to the departments. We don't meddle with that. Our assumption is that they have some kind of training in which they observe, or grade, an opportunity to deliver lectures or practice—in other words, not have total responsibility of that class. They will be trained hopefully for an entire two semesters, and then after that two-semester process they are no longer a rookie. But we do not structure the training. . . . We assume that they are doing the right thing.

And, in talking about the distribution of funds under the policy:

> If they need TAs to people 2000-level lab sections, then that's the place they should look to put an international immediately, before we need to buy them out. So that's another little nuance of it. But I don't know how many departments actively search other spots before they come over for their money. . . . In terms of the process of distributing that money, we just ask them how much and how many, that's all we ask. We do not verify. We do not double check.

The only sense the departments had that their activities were being monitored was through a requirement to submit a form to the provost's office on each graduate student funded under the NROR policy, a step that most considered *pro forma*. As one head commented,

> The only oversight I know is filling out those forms at the end of the semester. So we simply write down that we did this and we asked them to go through these lectures and the Instructional Effectiveness Program and just send it on over to the vice president's office. We never hear back one way or the other. That's the only oversight that there has been.

Despite the lack of monitoring, the departments achieved at least a minimal level of conformity with the policy. Those few departments to which the policy applied conscientiously undertook the altered teaching assignments and GTA training required by the policy. One department head noted that

> there are very few people that we have let out of the rookie program unless they have demonstrated with more than just a simple notation that they have taught or that they have some idea of what teaching is all about. We're not trying to save money by shunting them into the classroom by saying, "Oh, well, that is kind of like teaching." No, we haven't done that.

Another added,

> Some of these international students do very well. Nevertheless, we have abided by the policy and not used them in 1000-level courses. . . . We interpreted the policy as conservatively as we could and decided to apply it to all our classrooms, laboratories, and recitations.

Influences on Policy Implementation

In our analysis, we were able to identify several elements that influenced unit approaches to implementation of the policy. We found the level of sympathy with the public concerns that motivated institutional policy to have a minor impact. Of greater influence was whether the implementing units had previously attended to potential problems and solutions or had not. If they had, units took a more opportunistic approach to implementation and were able to address the choice opportunities inventively; if not, the departments were reactive to the new conditions the policy produced and approached implementation as compliance as defined at higher levels of the university hierarchy.

Public Concerns

In announcing the NROR policy, the president clearly stated that it was motivated by concerns among the general public of the state and particularly those who were or could become the parents of prospective students. President Halligan indicated as much in his memorandum announcing the policy on campus:

> Let me emphasize that the issue is not a matter of convincing the public that we have addressed the problem in the past or that the problem is not as serious as they perceive it to be. The public has heard that and is not convinced. We must do everything we can to change the perception and to immediately address the real problems that surface. (Halligan, memo, 1995, p. 1)

Thus, the policy was to symbolize to the people of the state the institution's commitment to undergraduate education and responsiveness to the expressed preferences of its students and others.

And the point was not lost on people inside the institution. One person told us the president "claims, and he tells the story all the time, that when he goes out and talks to the Moose Lodges and the Kiwanis Clubs and the Chambers of Commerce, and he tells them about the 'No-Rookies-on-Rookies' policy and he gets a standing ovation." The policy, in effect, signaled that the institution must be attentive to public concerns and imbued these concerns with some institutional authority. Therefore, reactions to the policy within the institution could be connected to beliefs about the legitimacy of those public concerns.

Many of the people we interviewed saw the need for and benefit from the policy. They recognized parents' desires to get the best education possible for their children. As one person put it, "Well, I appreciate what Dr. Halligan wanted to do in terms of politically reaching out to parents to make sure that they understand that OSU really cares about the kind of education people are getting." Another person added, "I can appreciate the fact that, if I was paying tuition, I would want my kid to have the best instruction, and presumably I want more than somebody that had just graduated the year before. So I understand all of that." As another person suggested, the positive feelings the policy inspired might be beneficial. "I think that any time President Halligan can get a standing ovation when he goes out and talks to groups around the state, that's a very positive thing for us."

However, several of the people we interviewed also dismissed the policy as little more than a public relations ploy, and they questioned the legitimacy of the concerns at the same time that they acknowledged the problem with the public image behind the policy. Comments such as, "The policy is not unreasonable, given the public outcry for having faculty teach and limiting the use of graduate TAs," and "the real problem the NROR policy was addressing was the public and legislative perception that faculty are not doing their jobs and are not accountable" have a suggestive undertone implying the outcry and perceptions were not justified. As discussed previously, most people we talked with did not think the public beliefs matched the reality of their departments. One person claimed that, in embracing the policy, the president "owned up to a problem that didn't exist," and he characterized the NROR policy as "silliness that wasted a lot of time."

In addition, those we talked with expressed concern about the apparent and unwarranted discrimination against international students that they believed the NROR policy embodied. For example, one commented:

> Well, my perspective is that we really need to be open to international students more. It's almost as if we bring them here then tell them they are not good enough to do a little teaching, and I think that is incorrect, at best. They are some of our finest students. They understand [my discipline] perhaps better than many of the American students do at that level. I don't want to say that they are smarter but they certainly have the dedication and discipline that some of the American students don't yet have.

Another added that the treatment of international students in the policy, "makes me very uncomfortable. I would accept any skill-based criteria, but this is strictly about country of origin." Respondents also felt that complaints about any language barrier were overstated and more often reflected cultural misunderstandings, rather than problems with the GTA's ability to speak English. Some also suggested that a problem with an international GTA's accent was an acceptable excuse for poor performance, but might not reflect the real reason a student was not doing well in a class. To underscore this, one person told us,

> Since we teach multiple sections of courses, it's easier to move the students out. Usually if students have a serious problem with the TA's language, I find those are rarely solved in the class, because the students already have a set attitude. But I will take students from two different sections, both claiming language problems, and I will switch sections between them. When I check back, they say, "Oh, I can understand that one. There's no problem now."

In fact, some went so far as to suggest that the public's focus on the issue reflected a parochialism that runs counter to the internationalization of all spheres of society today.

> The part that has bothered me most is the international students policy. I'm not in favor of that at all. I think undergraduate students use accents as an excuse and opportunity to say that they couldn't understand their teaching assistants, and I hate that we are encouraging that attitude. I think it encourages the xenophobia that you find out here. I've heard that same complaint from students not only in Oklahoma, but also in every place I have taught that is not located on a coast.

Another added, "It's a disadvantage if our students think that everyone is going to be like them. They need to kind of suck it up and be able to cope with other people from other cultures and other languages and things like that."

Thus, people we talked with generally discounted the public perspectives that the NROR policy addressed. Their acceptance of and compliance with the policy arose less from a perception of the legitimacy of the values and concerns that underlay the policy, but could be attributed more to either (a) structural factors such as the support for the policy within the institutional hierarchy (especially from a new and popular president), (b) the financial backing for the policy that prevented it from having a negative impact on graduate programs, and/or (c) the opportunity to address other issues and problems through the mechanisms of the policy.

Unit Contexts and Responses

Changes made by the implementing units varied substantially. The primary factor contributing to that variation was how each unit conceived of the NROR policy. In some cases, to facilitate implementation, units engaged in substantial redefinitions of the policy. They saw the policy as a vehicle for change. Other units appear to have mirrored policy definition and implementation from higher levels in the institutional hierarchy. They chose to comply with the policy through a literal interpretation of parameters and possibilities. Table 2 summarizes these aspects of the implementation processes in each of the academic units studied.

Opportunism. Clearly, College 1 was able to parlay the attention on freshmen retention as demonstrated by the NROR focus and financial backing into support for an initiative to create a freshmen seminar and insert it as a requirement in students' programs. According to one college administrator,

> We saw an opportunity to do some things we were planning to do anyway. So, we basically took advantage of the announcement and the charge to implement to do something that we had been working on, and that's our Freshmen Experience. . . . The "No-Rookies-on-Rookies" is, in some cases, looked at as a narrow concept, and in some sense, we've looked at it as a broad concept. We've looked at it as more than just not having a neophyte in the classroom in the freshmen year. We've looked at it more in terms of the total experience the undergraduates received in their freshmen year.

Staff of the college had been working on the project prior to the NROR announcement, but they realized quite early that the new policy created an impetus for getting their Freshmen Experience designs implemented.

Within College 2, a few departments demonstrated the same type of opportunism, but others did not. As one person observed, "I think there were not many departments that realized that this might benefit them. There were a couple that did, but most of them did not." Department A is one that took full advantage. As with College 1, the framework for seizing the opportunity had already been established. The study participant from Department A noted, "If you go back to 1990, we were doing changes in the [freshmen-level course] that helped. See, we were focused on changing [that course]. That was where we had the complaints."

Department A wanted to get away from having one course as the universal requirement for all students, and the departmental faculty had been developing alternative classes to better suit students' needs for at least three years. In their eyes, the problem was not the graduate assistant instructors but the introductory course itself that created dissatisfaction and needed changing. The NROR policy created the opportunity for the department to move more aggressively toward that goal and to develop a source of resources to help it happen. According to the departmental representative, "One nice thing was, this gave us a chance to look carefully at what we do at [the] entry [course] level. It prepared a fertile

TABLE 2

Interpretation and Organizational Responses to NROR Implementation

Unit	See a Problem	Legitimate Public Demand	Consider NROR Opportunity or Bother	Amount of Change	Decision-making Approach	Resources Received
College 1	Yes—Want contact with freshmen	Yes—Also concerned about effect of another college's classes on retention.	Opportunity	Substantial	Group—College administration and faculty	Moderate
Department A, College 2	Yes—Difficult freshmen course	Somewhat—Course, not TA teaching, the issue	Opportunity	Substantial	Group—Departmental administrators and the graduate committee	Significant
Department B, College 2	Yes–Problem was application of NROR program	Somewhat—Need to respond to the public	Bother	Substantial	Individual—Chair	Significant
Department C, College 2	Yes–Problem was application of NROR program	No	Bother	Substantial	Individual—Chair	Significant
Department D, College 2	No	No	Bother	Modest	Individual—Chair	Moderate
Department E, College 2	No	Somewhat—Understand parents' desires	Bother	Modest	Individual—graduate coordinator	Little
Department F, College 2	No	Doubtful	Bother	Modest	Individual—Chair	Little
Department G, College 2	No	No	Little concern	Little	Individual—Chair	Little

seed bed for what we've been doing with new courses. We told ourselves, 'Well, we better start these classes.'" While the department began offering new alternative freshmen-level classes, it now could afford to hire visiting faculty members, typically recent doctorates, to replace the GTAs in the standard freshmen-level course if they also increased the section size. "It's not directly the 'No-Rookies-on-Rookies' money, but it maybe made it possible because the 'No-Rookies-on-Rookies' is now paying for a set of TAs, so we have extra TA money that we can use."

Clearly, both College 1 and Department A of College 2 were opportunistic. They saw in the NROR policy the occasion to advance goals they had already begun to work toward. In College 1, beliefs about the necessity of intervening into the freshmen experiences of its majors led to reconsideration of their approaches to orientation and instruction for freshmen, and the NROR policy served as the vehicle promoting delivery of enhanced services to those students. In Department A, the recognition that one freshmen-level course could not serve the needs of all freshmen preceded the NROR policy. Both units were poised to redefine the policy in terms that fit their plans and ongoing efforts or to redefine their activities to fit the NROR initiative. In the process, they took advantage of the funds underwriting the new institutional policy and gained goodwill within the institutional hierarchy through engagement in the NROR program.

These two units also used collaborative decision-making processes to redefine the policy initiative and establish commitment and action to realize their new directions. They used groups of people to work on defining and preparing for their approach. According to our Department A informant, "We were changing the pedagogy in [the freshmen-level course]. . . . We would be the department administration and the graduate committee." And in College 1 an informant noted:

> For years I wanted to do something very different. So I came to the management team of this college and I said I think we need to reevaluate what we have done. I think we need to have contact with the freshmen year. So, brainstorming, we came up with the idea that we would try this team-like experiment for our entering freshmen.

Compliance. Departments B through G promoted different responses to the policy (see Table 2). The issues they addressed in their implementations were those of following the policy and addressing its implications for staffing classes. They did not see a problem in their own unit for which NROR provided a solution beyond the generic and obvious admission that it is probably better to have trained and experienced GTAs in freshmen-level classes. In effect, the departments were passive in defining the problems and reactive in their responses to them. They were "done to" and responded mechanically, rather than actively defining and creating responses.

In two departments, B and C, the problems and responses defined for them were more comprehensive and extreme than in the remaining four departments. These two departments were told to completely alter their approach to teaching their 1000-level courses after the dean's office determined that NROR, in their cases, meant that full-time faculty members rather than GTAs would teach the courses. This suggests some opportunistic use of the NROR policy by the dean's office staff, using the policy to affect specific changes when the conditions allowed. In the process, these two departments were able to use the extreme demands that policy implementation placed on them to get new resources in the form of additional faculty positions. As one department head explained it: "As a quid pro quo, the dean's office let the department fill a vacant position that had been left unfilled for a time." In addition, he was able to initiate "some discussions that will result in the transfer of a current faculty member to the department."

The remaining four departments, D through G, implemented the NROR policy in a relatively straightforward manner. They did not feel that the problem of rookies and international GTAs applied to their departments. "In [Department E], again we never had any graduate students teaching classes so, although there were complaints, we don't think we really had a problem." And, according to another department head, "We have had some problems in our classrooms with our teaching assistants, but they weren't necessarily rookies in the old days. Some of them were, but certainly not all of them, and I would certainly not say that all of the problems we had were the result of rookies." Nevertheless, they rearranged teaching assignments to implement the policy as defined by the offices of academic affairs and the arts and sciences dean, assured that they would get the resources to fund the graduate students shifted from teaching assignments, and developed some form of training for new graduate teaching assistants.

All of the departments that complied with the mandate and did not redefine the policy for their own plans and purposes also used decision-making processes that were less elaborate and participatory. They simply duplicated, within their own department, the top-down decision processes to which they were responding. Comments from three different interviews were: "I simply assigned everyone to teach [the freshmen-level course]. . . . I just proposed my solution to the faculty and they generally accepted it without a big flap;" "I usually handled most of that by myself;" and "I established the mentoring program for the graduate students and for the TAs, in particular." A single individual took responsibility for making decisions and implementing the necessary adjustments. These departments had not been working on ways to adjust their freshmen-level instruction, so they did not have different frameworks for considering the "No-Rookies-on-Rookies" issues, nor did they generate different frames when confronted with the implementation of NROR. They also did not consider there to be a problem and therefore did not see a reason to act in ways that might lead to more significant changes. Instead, individuals responded administratively to the implementation and procedural actions NROR required.

Organizationally, the interpretations of actions and reactions nicely fit the organized anarchy model (Cohen & March, 1974; March & Olsen, 1979; March, 1994). The policy, ambiguously defined upon inception, evolved multiple and diverse meanings and implications over time in different departments throughout the university. The NROR policy also served as a vehicle to identify and remedy related but distinct problems in different units. One college used it as impetus to create a freshmen orientation program, and the other college used this policy to remove *all* GTAs from freshmen-level courses in two departments. Some departments saw the policy as an opportunity to improve the freshmen-level classes of the department, to initiate new graduate student training, or to create new funding mechanisms for graduate students and new faculty positions. Other departments saw the policy as an unwelcomed problem to be dealt with—it resulted in more work for heads and different work for faculty but achieved only insignificant changes instructionally or academically.

In effect, the people in each implementing unit had to make sense of the new requirements and find some way of accommodating the conditions of the policy within their units. The people we talked to did not accept NROR as a symbol of admitting to unacceptable educational practices or of verifying public concerns about foreign students acting as GTAs. They generated alternative constructions for the meaning of the policy mandate. Some interpreted the policy as a public relations posture which they had to accept as a matter of hierarchical compliance. Others were able to redefine the NROR initiative to fit previously acknowledged needs or problems. To this latter set of implementors, NROR became an expression of the institution's broader concern for improvements in freshmen education and a conduit of resources for needs already acknowledged. These forms of meaning constructing, which required alternative definitions of demands and possibilities, had to include interchange that allowed the social aspects of sensemaking to operate. In other words, it required "rich media," the kind of direct personal contact that allows people to propose new meanings and negotiate them in order to settle on acceptable, new definitions (Daft & Lengel, 1986; Weick, 1995). The units that accomplished a proactive and opportunistic definition of the NROR policy had to operate more inclusive decision-making processes, including meetings in which people discussed the implications of the policy for their units and how best to respond to the policy.

Conclusions

The implementation of the NROR policy at OSU resulted in changes in freshmen-level instructional practices in the few units that fell under the formal definition of the policy. The scope and impact of the changes varied broadly. We have examined these changes and the factors influencing responses, including departments' reactions to the public concerns that motivated institutional leaders to design and adopt the NROR policy.

Institutional Responses

As a result of the NROR program, approaches to training rookie graduate students and to freshmen-level instruction changed. Some departments employed some form of training for GTAs while also removing them from direct responsibility for teaching 1000-level courses. Freshmen instruction changed most in four units: College 1 designed and implemented its Freshmen Experience as a new approach to orienting and retaining freshmen, Department A conceived new 1000-level courses, and Departments B and C shifted instruction to full-time faculty and increased class sizes.

Decision-making processes also differed across units in this study. Units reporting minor changes in response to the NROR policy made top down decisions. Department heads or other designated individuals devised solutions and implemented them with relatively low levels of input from faculty and students in their departments. In contrast, units reporting major changes employed committees and collaborative decision-making structures. Department heads and college administrators recognized certain concerns and were able to link them with the president's NROR policy. They engaged in garbage can decision making at its best (March & Olsen, 1979).

In effect, the institutional administration mediated the intent of the public through its definitions of the NROR policy within the institution; then the departments further mediated the institutional influences by how they accepted or refined the problem identified. The forms of implementation ranged from major changes producing very different educational experiences for students to nominal conformity in response to the specific directives that produced only slight differences in freshmen-level classes. The nature and impact of the implementation were substantially dependent upon the context, needs, capacity, and resourcefulness of the first-line implementors. In effect, the actions taken were several layers removed from the desires of the public, in whose name the policy was pursued institutionally. We cannot say to what extent the public would be content with the NROR implementation, if there were awareness of the narrow application of the policy created by the definitions of "freshmen," "rookie," and "international" or the modest changes it produced in some of the implementing departments. As so often happens in policy implementation, the countervailing influences on the implementors and the discretion they exercise result in organizational actions that reward the opportunistic and resourceful as much as the population the policy is intended to serve. However, as a symbol of institutional responsiveness to public concerns for undergraduate education, the NROR policy no doubt served its purpose.

Advice for Institutional Practice

Given the study findings and our conclusions, advice for those designing institutional responses to public concerns is clearly two-sided. We must acknowledge the practical necessity of the ambiguousness of the NROR initiative and the accomplishments of the unfettered implementation processes, conditions that could be equally productive in policymaking on other issues and in other types of institutions. As Yanow (1996) suggests, "To see ambiguous policy language as a problem to be solved in order to improve implementation chances is to ignore the reality of purposive ambiguity: it temporarily resolves conflicts and accommodates differences, allowing contending parties to legislate and move on to implementing actions" (p. 228). Policy in such cases becomes a symbol, with messages directed to both external and internal constituents. Externally, the policy demonstrates the institution's concern and willingness to take action to address a problem. Internally, it demonstrates the need to respond to external concerns. But beyond that, a loosely defined policy not only addresses the public concerns but also affords flexibility in implementation so that new activities can also address related concerns of internal constituents. For those charged with implementation, compliance enhanced by a creative linking to existing problems and goals can further unit development. In effect, a balance of external and internal concerns can be brought to bear on the policy issue.

In the case of the NROR policy, the president issued his policy mandate after limited consultation within the institution. That type of hierachical policymaking is not the norm in American research universities, often viewed as the exemplars of organized anarchy (Birnbaum, 1988). However, in this case, the specific policy is more accurately viewed as a symbol for an intent to push larger efforts to attract and retain undergraduate students and a demonstration of the need to act quickly toward that goal. In theoretical terms, the NROR policy presented ambiguous goals and engaged unclear technology, fundamental traits of garbage can decision-making processes. It was these aspects of the policy and the flexibility they permitted for implementers, along with the supporting funding, that allowed NROR to avoid crippling obstruction at the departmental level.

The offices of the provost and college dean, of course, had to provide further specifications to the policy. These served, in many instances as the guides for understanding the policy and as the recipes for departmental compliance. However, all the actors in hierarchical levels above the departments remained open to multiple interpretations of the policy. Even more important, they also left the monitoring and oversight of the policy thin and unobtrusive. With the exception of College 2's Departments B and C, higher-ups did not impinge on departmental creativity in responding to the presidential charge. Again applying the terminology of garbage can decision processes, they allowed various solutions to become attached to the NROR choice opportunity. As March (1994) has noted, garbage can processes do not require unrestricted choice opportunities and the free flow of problems. Rather, "the critical element of a garbage can process is that there are elements of temporal sorting. Linkages are formed, in part, because of simultaneity" (p. 205). In the case of NROR, both deans' offices and Department A were able to join already active problems or solutions to the policy and to promote activity and secure resources to support their decisions.

Whether he intended it or not, the president's policy, by allowing for flexibility in responses and rationale, used the characteristics of an organized anarchy to best advantage. The policy set the direction, communicated key values (attend to our public, improve retention, act quickly), provided necessary resources, and allowed unit decision makers to exercise their professional judgment and design responses appropriate to their contexts. We believe the policy was more externally directed and symbolic. Internally, the policy allowed decision processes to work in a variety of ways for the institution. The result, at a minimum, was restricting the countries of origin and providing prior training for people teaching 1000-level classes. However, this policy also created opportunities and provided the resources that allowed more robust solutions to other problems to be implemented within the institution.

References

Baier, V., March, J., & Saetren, H. (1986). Implementation and ambiguity. In J. March (Ed.), *Decisions and organizations* (pp. 150–166).Oxford: Basil Blackwell.

Ball, S. J. (1997, April). *Recreating policy through qualitative research: A trajectory analysis.* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Birnbaum, R. (1988). *How colleges work: The cybernetics of academic organization and leadership.* San Francisco: Jossey-Bass.

Bowe, R., & Ball, S J., with Gold, A. (1992). *Reforming education and changing schools: Case studies in policy sociology.* London: Routledge.

Cohen, M. D., & March, J. (1974). *Leadership and ambiguity: The American college president.* New York: McGraw-Hill.

Colleges expand efforts to help teaching assistants learn to teach. (1992, March 4). *Chronicle of Higher Education,* p. A-17, A-18.

Connelly, W. (1982, March). *Training GTAs to teach: Persona, personality, and policy.* Paper presented at the Annual Meeting of the Conference on College Composition and Communication, San Francisco. (ERIC Document Reproduction No. ED 213 046)

Daft, R., & Lengel, R. (1986). Organizational information requirements, media richness, and structural design. *Management Science, 32*(5), 554–571.

Ferman, B. (1990). When failure is success: Implementation and Madisonian government. In D. Palumbo & D. Calista (Eds.), *Implementation and the policy process* (pp. 38–50). New York: Greenwood.

Fischer, F., & Forester, J. (Eds.). (1987). *Confronting values in policy analysis.* Newbury Park, CA: Sage.

Fischer, F., & Forester, J. (Eds.). (1993). *The argumentative turn in policy analysis and planning.* Durham, NC: Duke University Press.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing.* Thousand Oaks, CA: Sage.

Linstead, S., Small, R., & Jeffcutt, P. (Eds.). (1996). *Understanding management.* Thousand Oaks, CA: Sage.

March, J. (1994). *A primer on decision making: How decisions happen.* New York: Free Press.

March, J., & Olsen, J. (Eds.). (1979). *Ambiguity and choice in organizations* (2nd ed.). Bergen, Norway: Universitetsforlaget.

Mazmanian, D. A., & Sabatier, P. A. (1983). *Implementation and public policy.* Glenview, IL: Scott, Foresman.

Mills, M. (1998). From coordinating board to campus: Implementation of a policy mandate on remedial education. *Journal of Higher Education, 69,* 672–697.

Palumbo, D., & Calista, D. (1990). Introduction: The relation of implementation research to policy outcomes. In D. Palumbo & D. Calista (Eds.), *Implementation and the policy process* (pp. 3–18). New York: Greenwood.

Pondy, L., Frost, P., Morgan, G., & Dandridge, T. (Eds.). (1983). *Organizational symbolism.* Greenwich, CT: JAI Press.

Proposed policy discriminatory. (1995, February 27). *Daily O'Collegian,* p. 4.

Weick, K. (1995). *Sensemaking in organizations.* Thousand Oaks, CA: Sage.

Yanow, D. (1996). *How does a policy mean?* Washington, DC: Georgetown University Press.

Janet G. Donald
D. Brian Denison

Quality Assessment of University Students

Student Perceptions of Quality Criteria

Quality assessment in higher education is of global interest; government and public demand for accountability from higher education institutions has steadily increased over the past decade (Brennan, Fedrowitz, Huber, & Shah, 1999). Given this demand, the need for ensuring the validity and utility of the assessment process has also increased. One issue of validity is whether the assessment responds to its audience, those who have mandated the assessment as well as those who will be affected by it. To be useful, the assessment must meet the needs of the people whom it is intended to benefit and aid the evaluated institution to make improvements. Quality assessment is frequently undertaken in response to external authorities who expect clear, ratified criteria to be used in the accountability process. If the assessment is to be beneficial, however, change must be effected within the institution. This means that administrators, faculty members, and students also need an understanding of the criteria that can guide and facilitate improvements in the way they function. Thus, to be both valid and useful, the approach taken to assessment must be aligned or consistent with the goals of the institution; assessment results must be credible and incorporable. The

This article is based on research funded by the Social Science and Humanities Research Council of Canada and by les Fonds pour la formation des chercheurs et aide à la recherche du Québec.

Janet G. Donald is professor of education at the Centre for University Teaching and Learning and the Department of Educational and Counselling Psychology, McGill University; D. Brian Denison is coordinator of student success at Champlain Regional College, Sherbrooke, QC.

Correspondence concerning this article should be addressed to the first author at 3700 McTavish St., Montreal, PQ, Canada, H3A 1Y2 or donald@education.mcgill.ca.

The Journal of Higher Education, Vol. 72, No. 4 (July/August 2001)
Copyright © 2001 by The Ohio State University

criteria on which an assessment is based must fit the institution's goals and priorities as well as being understood by the larger community.

Indicators of student quality have been used regularly as measures in the evaluation of postsecondary institutions. However, because different types of stakeholders in higher education—the public, administration, faculty, or students—have their own perspectives and goals, they assign different values to criteria of quality. The public wants students to graduate with general abilities and emphasizes criteria such as communication skills (Cave & Hanney, 1992). University administrators are expected to show that resources are being used efficiently and effectively; for them, student completion of program requirements is an important criterion (Nadeau, Donald, & Konrad, 1992). Faculty view the primary obligation of the university to students to be the development of intellectual independence (Baird, 1988; Barzun, 1993) and hence focus on criteria such as the ability to think critically (Barnett, 1988; Trice & Dey, 1997). Students, meanwhile, are increasingly preoccupied with career concerns; they value criteria such as the ability to get a job (Dey, Astin, & Korn, 1991). To come to an understanding of higher education goals and criteria for judging success in meeting them requires an analysis of the value attached to them by different stakeholders.

The issue of aligning the approach taken to assessment with the goals of the institution is equally complex. The most common approaches to quality assessment are based on reputation and resources (Astin, 1985). Reputation is a global assessment of the perceived status or excellence of an institution or program, typically measured by asking knowledgeable experts to rate the institution or program. For example, presidents of research universities might rate other research universities for excellence (Cave & Hanney, 1992). The institutions or programs are then ranked according to their averaged ratings. The resource approach to quality uses input measures of faculty and student quality and physical and fiscal resources (endowments, per-student expenditures). Within the resource approach, one of the most frequently used measures of quality is the academic ability of students at entry to college or university. A third approach uses performance indicators or global outputs to define quality—indices such as program or degree completion rates, the proportion of undergraduates admitted to graduate education, or alumni satisfaction ratings.

The flaw in these approaches is that they do not suggest how improvements in quality could be made. Arguing that these approaches to quality are counterproductive, Astin (1985) recommended that quality be considered in terms of talent development. The criteria for quality in Astin's approach are grounded on an institution's ability to affect its students and faculty favorably; for example, by enhancing the knowledge and personal development of students and the scholarly and pedagogical ability and productivity of faculty members. An essential component of this approach is simultaneously considering inputs, environments, and outcomes (Astin, 1993). The effective institution is one that adapts programs and policies to improve students' educational experience.

Programs consistent with this approach frequently focus on the assessment of student progress. For example, Alverno College developed an abilities-based curriculum with regular assessment of the level of ability shown by individual students (Mentkowski & Loacker, 1985). Another widely recognized assessment program uses a series of tests of student competencies (Banta, 1989). In these programs, student performance at entry to, during, and exit from college is compared to establish the extent to which student competencies increase. Student outcomes can be differentiated in terms of the *kind* of outcome and the *temporal context* of the outcome (Astin, 1977, 1993; Ewell, 1991; Pascarella & Terenzini, 1991; Terenzini, 1989). The kinds of outcomes measured include: (a) knowledge, (b) skills—basic, higher order, and career related, (c) attitudes and values, and (d) behavioral outcomes—what students do during and after college or university. Recognizing temporal context means that learning and development are assessed at entry and during the course of studies and that outcomes at one point in time become inputs at the next. Thus, it becomes possible to make distinctions between short-, intermediate-, and long-term outcomes.

In the measurement of academic progress, indicators of the nature and extent of interactions between the student and the educational environment are important intervening variables. One measure of interaction, involvement, the extent to which students invest physical and psychological energy in their educational experience, has emerged as a major construct in ensuring positive outcomes (Astin, 1984, 1996; Kuh, Schuh, Whitt, & Associates, 1991; Pace, 1988, 1990; Pascarella & Terenzini, 1991; Willis, 1993). Motivation, self-efficacy and self-regulation are other indices of interactive effects (Donald, 1997, 1999; King & Baxter Magolda, 1996; Pintrich, 1995; Pintrich & Garcia, 1994). Pintrich notes that self-regulated learners monitor and attempt to control not only their behavior, but also their motivation and affect; students are active agents in their learning. The assessment literature thus suggests that criteria of student quality should not be limited to input and output measures, should acknowledge the influence of environmental and attitudinal factors, and will not necessarily be discrete or independent of each other but rather are likely to be interactive.

Students are a central focus in assessments of educational quality.

What is less frequently acknowledged is that they are also major stakeholders in postsecondary education. To date, quality criteria have reflected administrators' or faculty priorities. As both the subjects of assessment and stakeholders, we argue that students and their perceptions of quality criteria need to be incorporated into the assessment process. For example, students have a different perception of grades, a central component of assessment, than do professors (Goulden & Griffin, 1995). Whereas faculty focus on the role of grades as feedback, students see grading as including a gatekeeping function. They are thus likely to assign a different value to grades than faculty do. In spite of differences between students' and other stakeholders' views, students' perceptions are rarely examined and so cannot be taken into account in the assessment process.

Postsecondary institutions with the greatest educational impact are those with clear and consistent educational goals that are shared by faculty and students (Chickering & Reisser, 1993; Schroeder & Hurst, 1996). The importance students accord to particular goals shapes the activities in which they engage and the degree of effort they expend in those activities (Cantor & Langston, 1989; Cantor et al., 1991; Volet, 1997; Volet & Lawrence, 1990). However, there is evidence of increasing divergence in the educational goals of faculty and students. For example, faculty teaching goals have remained relatively stable over the past two decades, with highest priority given to facilitating students' intellectual development (Trice & Dey, 1997). In contrast, students are increasingly interested in obtaining practical training for employment (Dey et al., 1991).

Left unaddressed, the presence of divergent goals between faculty and students is likely to lead to increased levels of frustration and dissatisfaction. Faculty express dissatisfaction when they perceive that students are not committed to intellectual pursuits. Students manifest disappointment with the academic advising they receive, because it does not respond to their needs (Alexitch, 1997; Herndon, Kaiser, & Creamer, 1996). Criteria for student quality reflect a range of educational goals. It is therefore important that assessors understand the perspective of students in order to identify first, potential areas of conflicting interpretation that might diminish the validity of assessment and second, steps that should be taken to improve the institutional environment.

Given these issues, the purpose of this study was to examine students' perceptions of quality criteria previously identified by a broad range of stakeholders in a national study of criteria and indicators of quality in postsecondary education (Donald & Denison, 1993; Nadeau et al., 1992). The stakeholders were governors of university and college boards, administrators, faculty, students, and members of the larger community. They independently identified criteria of quality in universities and colleges that they considered to be important, then verified the criteria over three rounds of a Delphi procedure (Linstone & Turoff, 1975). The 25 criteria for the quality of students that emerged from this study reflected the perspective of the broad postsecondary education community, but not specifically students' own views.

Based on our experience with the Delphi study and the extant research on assessment, we had four objectives. First, given the relatively limited number of students in the Delphi study and their disparate backgrounds, we sought to determine the extent to which a representative sample of students in the same postsecondary institution would view the criteria as important indices of student quality. A single institution was selected in order to replicate as closely as possible the conditions in which quality assessment would be undertaken. Second, we sought insight into the relationships between the 25 criteria. Would students link certain criteria together and discriminate between others? Would indices of involvement or self-regulation be associated with cognitive outcomes? Third, we sought to determine whether students perceive the importance of these criteria as constant across their undergraduate years, or variable, depending upon the particular point in an undergraduate career when student quality is assessed.

Finally, students' postsecondary experiences and outcomes have been shown to vary according to their gender (Baxter Magolda, 1990; Donaldson & Dixon, 1995; Meyer, Dunne, & Richardson, 1994) and program of study (Cashin & Downey, 1995; Pascarella & Terenzini, 1991). Differences in students' values and attitudes have been found based on cohort, that is, their year of study or level of postsecondary experience (Constantinople, 1967; Feldman & Newcomb, 1969; Martin, Ramsden, & Bowden, 1989). We therefore sought to determine the effects of gender, program, and cohort on student perceptions of student quality.

Method

Sample

Students selected for the study were enrolled during the fall semester of 1994 in the four major undergraduate programs in the university, in arts and science (in English literature, physics, and psychology courses), education, and engineering at a research university (Table 1). The sample of 400 students was representative of the undergraduate student population in terms of entering average (85%) and gender composition (58% female) (McGill University, 1996). Students ranged in age from 18 to 24 with a modal age (41%) of 19 years. For 25% of the students, this was their first postsecondary experience. Another 41% had graduated from high school and had participated in a pre-university college program—a normal educational route in the province—and were in their first year of university. The remaining 34% were in their second or third year of university. Thus within a representative sample consisting of students from pure and professional programs, we had three tiers of undergraduates with increasing postsecondary experience.

Procedure

A questionnaire designed to examine student learning at the university was administered to students in class in the third week of the fall semester. Students were informed that their participation was being solicited to achieve a better understanding of students' university experience, and to improve university teaching and learning. In one section of the questionnaire, students were presented with the set of 25 criteria accompanied by an explanation that a study of indicators of quality in higher education had suggested them as important for students at different stages during university—at entry to university, while pursuing a degree, and upon graduation. Students were asked to use a 5-point response scale (1 = not at all important, 2 = somewhat important, 3 = important, 4 = quite important, 5 = extremely important) to indicate how important they felt each criterion was for evaluating the quality of a student at each point in time. Thus, each criterion was to be given three importance ratings. Because the procedures for data analysis are specific to each research objective, they are presented at the beginning of the relevant section in the results.

TABLE 1
Composition of Sample by Program of Study and Gender

	Program of Study									
	Arts		Science		Education		Engineering		Total	
Gender	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Male	42	35.3	26	44.8	13	13.0	86	69.9	167	41.8
Female	77	64.7	32	55.2	87	87.0	37	30.1	233	58.3
Total	119		58		100		123		400	

Results

Comparison with Other Stakeholders

We first wanted to assess the consistency between the student ratings and those made by the broad range of stakeholders in the Delphi study. Because our sample consisted of university students, only the results from those respondents in the original study who specifically indicated they were rating the criteria in reference to university students ($N = 93$) were used for comparative purposes (Donald, 1990). In contrast to the student ratings, ratings in the Delphi study (a) had been made without reference to a particular point in a student's academic career, and (b) utilized a 6-point response scale (0 = of no importance, 5 = of very high importance). To obtain a general rating of importance that could be used for comparison with the results of the Delphi study, a composite of the three sets of students' ratings was made by averaging across time, thus decontextualizing the ratings. The criteria were then ranked for each group based on mean importance ratings (Table 2).

In the composite ratings, most of the criteria (22/25) were quite important ($3.50 \leq M \leq 4.49$), and the remaining three were important ($2.50 \leq M \leq 3.49$). No criterion was extremely important ($M \geq 4.50$). Three criteria in particular—commitment to learning; ability to analyze, synthesize, and think critically; and general academic preparedness—were among the most important criteria for both Delphi study stakeholders and students. The Spearman correlation between students' and Delphi study ranks of importance was 0.71 ($p = 0.000$). We could therefore infer that a representative group of students in this institution found the national criteria for student quality important and were in accord with the broader stakeholder population. More revealing, the consistency between the two groups in terms of which criteria were considered most important supported the relevance of generally accepted criteria for students—whether concerning their preparation, motivation, or thinking ability.

Relationships Among the Criteria

Would students perceive some criteria to be related? A principal components analysis of the composite ratings utilizing a varimax rotation extracted five factors, accounting for 57.3% of the variance (Table 3). The first factor encompassed 11 of the 25 criteria, combining cognitive and affective student characteristics. Loading most strongly on this factor, labeled generic skills and abilities, were (a) openness and flexibility in learning, (b) independence in learning, (c) a sense of responsibility, and (d) the ability to analyze, synthesize, and think critically. Thus stu-

TABLE 2

Comparison of Student Composite Ratings and Rankings with Stakeholder Results

Criterion	Student Composite *M*	*SD*	Rank	Stakeholders Rank
General academic preparedness	4.16	0.66	3.5	3
Secondary school preparation	3.12	1.01	25	8
Preparedness for a specific program	3.73	0.77	18	24
Breadth of life experience	3.55	0.88	22	23
Basic communication skills	3.99	0.74	12	4.5
Basic mathematical competency	3.49	1.00	23	16
Intelligence	4.00	0.82	11	8
Commitment to learning	4.33	0.68	1	1.5
Clarity of student's educational and career goals	3.62	0.86	21	22
Competence in second language	3.15	1.07	24	25
Sense of responsibility	4.11	0.76	5	12
Openness and flexibility	3.91	0.82	14	16
Independence in learning	4.09	0.75	6	8
Ability to analyze, synthesize, and think critically	4.16	0.66	3.5	1.5
Ability to interact with others	3.94	0.78	13	12
Effective study skills and habits	4.01	0.68	10	4.5
Moral and ethical reasoning	3.68	0.95	20	18.5
Personal student development	3.74	0.82	17	20
Self-confidence	4.09	0.78	7	12
Academic performance/achievement in courses	4.02	0.75	9	16
Completion of program requirements	4.24	0.75	2	12
Expertise at end of program	3.90	0.80	16	12
Ability to get a job	3.68	0.89	19	21
Performance on the job	3.91	0.87	15	18.5
Commitment to lifelong learning	4.04	0.92	8	6

NOTE: Composite ratings are based on the average of the importance ratings assigned to each of the three temporal contexts (at entry to university, during studies, and upon graduation) on a 5-point scale (1 = *not at all important*, 5 = *extremely important*).

dents associated indices of involvement and self-regulation with thinking. The second factor, labeled academic performance, included (a) academic performance or achievement in courses, (b) completion of program requirements, and (c) intelligence. A third factor described employment competence and consisted of (a) performance on the job, (b) ability to get a job, and (c) expertise at end of program. Basic mathematical competency and competence in a second language loaded together to form a fourth factor, specific skills. The fifth factor, academic preparedness, consisted of (a) secondary school preparation, (b) general academic preparedness, and (c) preparedness for a specific program. Three criteria—clarity of educational and career goals, commitment to lifelong learning, and breadth of life experience—loaded at levels lower than 0.50 across several factors, suggesting a more general relationship to student quality.

TABLE 3
Principal Components Analysis of Composite Student Criterion Ratings

Component and Criterion	Loading
Generic skills and abilities (21.3% of variance)	
Openness and flexibility in learning	0.76
Independence in learning	0.71
Sense of responsibility	0.69
Ability to analyze, synthesize, and think critically	0.68
Ability to interact with others	0.62
Moral and ethical reasoning	0.60
Commitment to learning	0.59
Self-confidence	0.57
Personal student development	0.57
Effective study skills and habits[a]	0.55
Basic communication skills	0.51
Academic performance (10.3% of variance)	
Academic achievement / performance in courses	0.77
Completion of program requirements	0.65
Intelligence	0.55
Effective study skills and habits[a]	0.50
Employment competence (10.1% of variance)	
Performance on the job	0.79
Ability to get a job	0.76
Expertise at end of program	0.58
Specific skills (8.8% of variance)	
Competency in second language	0.66
Basic mathematical competency	0.63
Academic preparedness (6.8% of variance)	
Secondary school preparation	0.76
General academic preparedness	0.63
Preparedness for a specific program	0.57

[a] Criterion loading ≥ 0.50 on two principal components

In reviewing these factors, we find that student perceptions of the criteria are consistent with previous research results on input and output measures. The criteria grouped as academic preparedness are those commonly considered academic inputs in the research literature, while the employment competence criteria are consistent with commonly reported output measures. The most important result of this analysis is, however, that the generic skills and abilities factor explains a large percentage of the variance and links students' personal abilities and attitudes with their cognitive abilities and skills. Students' perceptions of their abilities are thus more holistic than taxonomies of student outcomes would suggest (e.g., Pascarella & Terenzini, 1991). In addition, students view student quality in more comprehensive terms than do faculty (e.g., Barnett, 1988).

Influences on Student Perceptions of Student Quality

Four potential influences on students' ratings (temporal context, gender, program of study, and cohort) were investigated by means of repeated measures MANOVAs. To minimize the likelihood of spurious results, a Bonferroni adjustment was used to establish the critical probability level for determining statistical significance. As each repeated measures MANOVA entailed 15 statistical tests to evaluate main and interaction effects, the use of $p = 0.006$ as the critical level maintained an overall α of 0.09 for the family of 15 tests. Thus, we remained within the 0.10 overall level for directional hypotheses while maintaining a relatively high critical level ($p < 0.01$) for each statistical analysis. This significance level allowed for the further analysis of quality criteria considered important in the literature (e.g., mathematical competency, career goals).

Importance of criteria within temporal context. Students had been asked in the questionnaire to rate each criterion at entry to university, during studies, and upon graduation. Results thus reflect ratings given at one point in time but projected for the three temporal contexts. As shown in Table 4, students considered the 25 criteria to be important measures of student quality in each temporal context ($M \geq 2.50$). At entry to university, 15 criteria were deemed quite important ($3.50 \leq M \leq 4.49$). Several of these criteria (general academic preparedness, secondary school preparation, basic communication skills, and intelligence) are commonly used for making admissions decisions. However, students also considered a variety of other criteria to be quite important at entry to university: a sense of responsibility, openness and flexibility, independence in learning, the ability to interact with others, self-confidence, and commitment to lifelong learning. These findings lend support to the argument that admissions decisions should be based on more than students' having the necessary course prerequisites and evidence of academic aptitude. In particular, student commitment to learning is the most highly supported criterion of quality at entry to university, suggesting that some measure of commitment be included in a student's admissions dossier.

During studies, the number of criteria considered quite important increased from 15 to 23, and the remaining two criteria (secondary school preparation and competence in a second language) were deemed important as measures of student quality. The results for several of the criteria appear counterintuitive. For example, one would think of completion of program requirements as something to assess at the end of a program of study, yet students deemed it quite important during studies. They ap-

TABLE 4
Importance Ratings of Criteria in Different Temporal Contexts

	Temporal Context					
	At Entry		During Studies		On Graduation	
Criterion	*M*	*SD*	*M*	*SD*	*M*	*SD*
General academic preparedness	4.10	0.91	4.23	0.75	4.14	0.99
Secondary school preparation	3.82	1.09	3.02	1.21	2.53	1.35
Preparedness for a specific program	3.25	1.18	3.91	0.89	4.04	1.11
Breadth of life experience	3.11	1.14	3.53	1.00	4.02	0.95
Basic communication skills	3.62	0.95	3.94	0.85	4.40	0.82
Basic mathematical competency	3.41	1.06	3.52	1.07	3.53	1.10
Intelligence	3.87	0.91	4.01	0.88	4.12	0.87
Commitment to learning	4.25	0.83	4.48	0.68	4.26	0.87
Clarity of student's educational & career goals	2.90	1.19	3.73	0.99	4.23	0.96
Competence in second language	2.88	1.15	3.03	1.15	3.53	1.20
Sense of responsibility	3.76	0.98	4.12	0.84	4.45	0.76
Openness and flexibility	3.64	1.01	3.89	0.93	4.21	0.87
Independence in learning	3.78	0.99	4.19	0.80	4.32	0.81
Ability to analyze, synthesize, think critically	3.70	1.01	4.23	0.72	4.54	0.68
Ability to interact with others	3.60	1.00	3.92	0.89	4.30	0.84
Effective study skills and habits	3.81	0.91	4.31	0.73	3.93	0.99
Moral and ethical reasoning	3.39	1.08	3.69	1.01	3.97	1.02
Personal student development	3.49	0.96	3.85	0.86	3.89	1.00
Self-confidence	3.79	0.98	4.06	0.87	4.40	0.77
Acad. performance/achievement in courses	3.86	0.92	4.17	0.78	4.02	0.99
Completion of program requirements	3.98	1.08	4.34	0.82	4.41	0.90
Expertise at end of program	3.33	1.22	3.89	0.96	4.50	0.76
Ability to get a job	3.00	1.26	3.52	1.16	4.53	0.86
Performance on the job	3.36	1.27	3.74	1.15	4.63	0.67
Commitment to lifelong learning	3.79	1.12	4.03	0.98	4.30	0.91

peared to perceive completion of program requirements as an ongoing process, something to which they must attend throughout their studies. Both the ability to get a job and performance on the job were deemed quite important as criteria of student quality during studies. Students in engineering perceived the ability to get a job to be more important during studies than did arts students. However, students in all four programs considered performance on the job to be quite important during studies, reflecting the career and financial concerns of students.

Upon graduation, many criteria were given even higher ratings. Four were thought extremely important by students, in order: (a) performance on the job; (b) the ability to analyze, synthesize, and think critically; (c) the ability to get a job; and (d) expertise at end of program. Three of these criteria are related to employment competence and thus consistent with the career concerns of students. The fourth, however, reflects the type of intellectual outcome emphasized by faculty. Thus faculty and students are in accord about the relative importance of thinking abilities upon graduation.

Differences in importance of criteria across temporal context. For all but one of the 25 criteria, repeated measures MANOVAs revealed that temporal context had a significant main effect on importance ratings, the exception being general academic preparedness (Table 5). In order to gain insight into the temporal effects, post hoc multiple comparisons between the temporal contexts were conducted for each criterion by means of paired t-tests. Thus, ratings made for entry to university were compared with those for during studies and upon graduation; similarly, ratings for during studies were compared with those for upon graduation. For each family of three post hoc comparisons, an α of 0.05 was maintained by setting the critical level for statistical significance at $p = 0.018$.

Four patterns of difference in importance ratings were found: (a) decreasing, (b) peaking, (c) plateauing, and (d) increasing (Figure 1). One criterion displayed a decreasing pattern—secondary school preparation; students gave importance ratings that were (a) significantly lower during studies than at entry, and (b) significantly lower again upon graduation than during studies. A significant Temporal Context x Cohort interaction effect was also found for this criterion, Wilks' $\Lambda = 0.948$, $F(4, 750) = 5.05$, $p = 0.001$. Post hoc analyses revealed that students entering from college rated the importance of secondary school preparation significantly lower. This is understandable, because their acceptance into university depended upon their college preparation rather than their high-school preparation. This interaction effect did not, however, mitigate the main effect of temporal context; for each cohort, ratings on the importance of secondary school preparation decreased significantly from entry to during studies to graduation.

The second pattern, peaking importance, was characterized by (a) significantly higher importance ratings during studies than at entry, and (b) significantly lower ratings upon graduation than during studies. Four criteria displayed this pattern: (a) general academic preparedness, (b) commitment to learning, (c) effective study skills and habits, and (d) academic performance or achievement in courses. According to the students, these criteria are more related to success during one's studies than at entry or upon conclusion of an academic program. Again, certain results are counterintuitive. Students appear to consider academic preparation specific to their studies, thus giving it greater importance during studies than at entry, in contrast to the resource approach. The same pattern for academic performance or achievement testifies to students' specific rather than general orientation to their studies.

TABLE 5

Effects of Temporal Context on Ratings of Criterion Importance for Assessing Student Quality

	Main Effect			Post-hoc Contrasts			
Criterion	Wilks' Λ	*F*	*p*	Contrast	*t*	*p*	Conclusion
General academic preparedness	0.975	4.79	0.009				nsd
Secondary school preparation	0.669	92.84	0.000	ENT vs. DUR	14.05	0.000	ENT > DUR
				DUR vs. GRAD	11.34	0.000	DUR > GRAD
				GRAD vs. ENT	17.30	0.000	ENT > GRAD
Preparedness for a specific program	0.821	40.85	0.000	ENT vs. DUR	11.72	0.000	DUR > ENT
				DUR vs. GRAD	2.33	0.020	nsd
				GRAD vs. ENT	10.00	0.000	GRAD > ENT
Breadth of life experience	0.757	60.30	0.000	ENT vs. DUR	9.22	0.000	DUR > ENT
				DUR vs. GRAD	12.78	0.000	GRAD > DUR
				GRAD vs. ENT	16.17	0.000	GRAD > ENT
Basic communication skills	0.738	66.57	0.000	ENT vs. DUR	9.03	0.000	DUR > ENT
				DUR vs. GRAD	12.83	0.000	GRAD > DUR
				GRAD vs. ENT	16.19	0.000	GRAD > ENT
Basic mathematical competency	0.970	5.74	0.004	ENT vs. DUR	3.91	0.000	DUR > ENT
				DUR vs. GRAD	0.33	0.744	nsd
				GRAD vs. ENT	3.04	0.003	GRAD > ENT
Intelligence	0.943	11.29	0.000	ENT vs. DUR	4.83	0.000	DUR > ENT
				DUR vs. GRAD	5.11	0.000	GRAD > DUR
				GRAD vs. ENT	6.98	0.000	GRAD > ENT
Commitment to learning	0.877	26.22	0.000	ENT vs. DUR	7.96	0.000	DUR > ENT
				DUR vs. GRAD	5.85	0.000	DUR > GRAD
				GRAD vs. ENT	0.29	0.769	nsd
Clarity of student's educational and career goals	0.566	143.98	0.000	ENT vs. DUR	18.96	0.000	DUR > ENT
				DUR vs. GRAD	10.96	0.000	GRAD > DUR
				GRAD vs. ENT	20.46	0.000	GRAD > ENT
Competence in a second language	0.805	45.52	0.000	ENT vs. DUR	5.59	0.000	DUR > ENT
				DUR vs. GRAD	11.42	0.000	GRAD > DUR
				GRAD vs. ENT	13.48	0.000	GRAD > ENT

TABLE 5 (*Continued*)

	Main Effect			Post-hoc Contrasts			
Criterion	Wilks' Λ	*F*	*p*	Contrast	*t*	*p*	Conclusion
Sense of responsibility	0.770	56.07	0.000	ENT vs. DUR	11.03	0.000	DUR > ENT
				DUR vs. GRAD	11.57	0.000	GRAD > DUR
				GRAD vs. ENT	15.35	0.000	GRAD > ENT
Openness and flexibility	0.834	37.32	0.000	ENT vs. DUR	7.86	0.000	DUR > ENT
				DUR vs. GRAD	8.80	0.000	GRAD > DUR
				GRAD vs. ENT	12.39	0.000	GRAD > ENT
Independence in learning	0.827	39.11	0.000	ENT vs. DUR	11.36	0.000	DUR > ENT
				DUR vs. GRAD	3.81	0.000	GRAD > DUR
				GRAD vs. ENT	11.64	0.000	GRAD > ENT
Ability to analyze, synthesize, and think critically	0.728	69.94	0.000	ENT vs. DUR	14.49	0.000	DUR > ENT
				DUR vs. GRAD	9.10	0.000	GRAD > DUR
				GRAD vs. ENT	16.41	0.000	GRAD > ENT
Ability to interact with others	0.799	47.13	0.000	ENT vs. DUR	9.54	0.000	DUR > ENT
				DUR vs. GRAD	10.51	0.000	GRAD > DUR
				GRAD vs. ENT	13.94	0.000	GRAD > ENT
Effective study skills and habits	0.750	62.54	0.000	ENT vs. DUR	13.27	0.000	DUR > ENT
				DUR vs. GRAD	7.76	0.000	DUR > GRAD
				GRAD vs. ENT	2.01	0.046	nsd
Moral and ethical reasoning	0.836	36.75	0.000	ENT vs. DUR	9.99	0.000	DUR > ENT
				DUR vs. GRAD	9.11	0.000	GRAD > DUR
				GRAD vs. ENT	12.87	0.000	GRAD > ENT
Personal student development	0.865	29.27	0.000	ENT vs. DUR	10.48	0.000	DUR > ENT
				DUR vs. GRAD	1.33	0.185	nsd
				GRAD vs. ENT	8.26	0.000	GRAD > ENT
Self-confidence	0.774	54.73	0.000	ENT vs. DUR	9.22	0.000	DUR > ENT
				DUR vs. GRAD	10.96	0.000	GRAD > DUR
				GRAD vs. ENT	14.20	0.000	GRAD > ENT
Academic achievement/performance in courses	0.890	23.06	0.000	ENT vs. DUR	9.26	0.000	DUR > ENT
				DUR vs. GRAD	3.77	0.000	DUR > GRAD
				GRAD vs. ENT	2.89	0.004	GRAD > ENT

TABLE 5 (*Continued*)

Criterion	Main Effect			Post-hoc Contrasts			
	Wilks' Λ	*F*	*p*	Contrast	*t*	*p*	Conclusion
Completion of program requirements	0.871	27.89	0.000	ENT vs. DUR	9.57	0.000	DUR > ENT
				DUR vs. GRAD	1.53	0.128	nsd
				GRAD vs. ENT	7.00	0.000	GRAD > ENT
Expertise at end of program	0.654	99.20	0.000	ENT vs. DUR	13.00	0.000	DUR > ENT
				DUR vs. GRAD	13.47	0.000	GRAD > DUR
				GRAD vs. ENT	18.38	0.000	GRAD > ENT
Ability to get a job	0.577	137.71	0.000	ENT vs. DUR	13.10	0.000	DUR > ENT
				DUR vs. GRAD	17.55	0.000	GRAD > DUR
				GRAD vs. ENT	22.28	0.000	GRAD > ENT
Performance on the job	0.642	104.75	0.000	ENT vs. DUR	10.48	0.000	DUR > ENT
				DUR vs. GRAD	16.36	0.000	GRAD > DUR
				GRAD vs. ENT	19.80	0.000	GRAD > ENT
Commitment to lifelong learning	0.818	41.66	0.000	ENT vs. DUR	7.99	0.000	DUR > ENT
				DUR vs. GRAD	7.94	0.000	GRAD > DUR
				GRAD vs. ENT	11.54	0.000	GRAD > ENT

NOTE: For each criterion, the critical level for statistical significance of main and interaction effects was set at $p \leq 0.006$, thereby maintaining an experiment-wise α of 0.09. Degrees of freedom for Wilks' Λ $F = 2, 375$. Degrees of freedom for post hoc paired *t*-tests = 399, with critical $p \leq 0.018$, for an α of 0.05 for each set of 3.

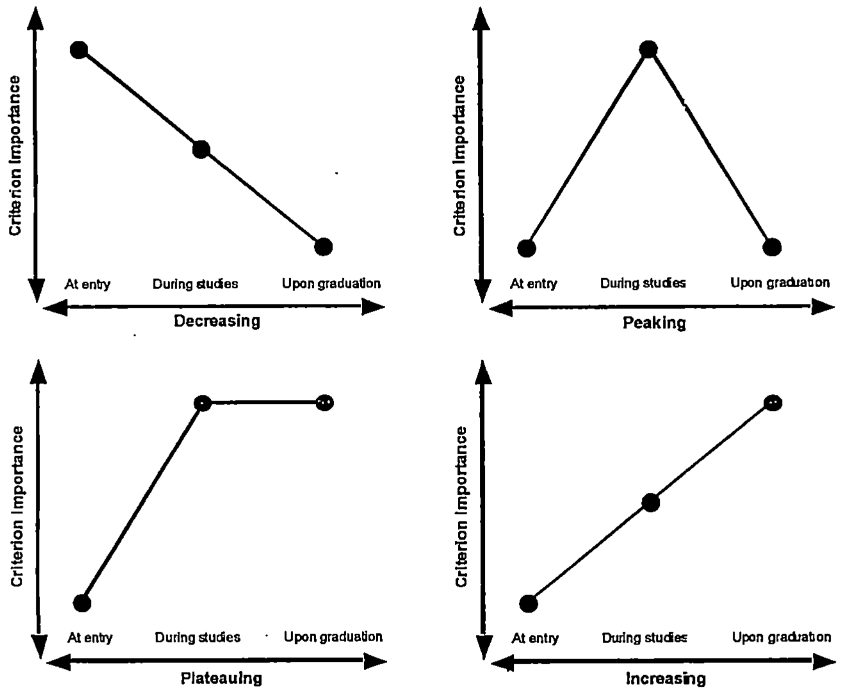


FIG. 1. Patterns of Change in Perceived Criterion Importance Across Temporal Contexts

The third pattern, plateauing importance, was characterized by (a) importance ratings that were significantly higher during studies than at entry, and (b) no significant differences between during studies and upon graduation contexts. Three criteria displayed this pattern: (a) basic mathematical competency, (b) personal student development, and (c) completion of program requirements. Again, a specific orientation on the part of students appears in their not giving significantly greater importance to completion of program requirements at graduation.

The final pattern, increasing importance, was characterized by (a) importance ratings during studies that were significantly higher than at entry and (b) importance ratings upon graduation significantly higher than during studies. The majority of criteria (17 of 25) displayed this pattern. It would appear that each individual step in the program is a point for students to review their progress. The large number of criteria increasing in importance attests to the student perception that outcomes are more important than what happens during studies, which in turn is more important than inputs.

The effects of gender, program of study and cohort on importance. In marked contrast to the pervasiveness of temporal context effects, only 5 of the 25 criteria were found to have significant gender, program of study, or cohort effects on the perceived importance of the criteria (Table 6). The only criterion to exhibit a significant gender effect unaccompanied by an interaction effect was general academic preparedness. Female students rated it significantly higher than male students did at entry and during studies.

Program of study influenced the perceived importance of (a) sense of responsibility and (b) clarity of student's educational and career goals. Education students rated the importance of a sense of responsibility more highly in all three temporal contexts than did arts students. Education students are possibly cognizant that their studies are preparing them to directly affect other people and therefore see responsibility as more pertinent and value it more than do arts students. Clarity of educational and career goals showed differences, primarily between arts students and those in professional programs. Education and engineering students rated the importance of goal clarity more highly at entry than did arts students. Education students also rated the criterion more highly at graduation than students in arts. A reasonable explanation for these findings is that students in professional programs require greater goal clarity than is the case in arts; establishing educational and career goals may be part of the agenda for many arts students. For those arts students who do not have clear educational and career goals, the findings raise the question of how this might affect their academic involvement or their ability to self-regulate.

Cohort affected importance ratings for two criteria: (a) secondary school preparation and (b) clarity of educational and career goals. Students entering from high school rated secondary school preparation significantly higher at entry than did students entering from college, as might be expected. Upper-year students also rated the criterion more highly at entry than did students entering from college. This might appear counterintuitive, except that a proportion of upper-year students had entered university directly from high school. For students entering from college, we could surmise that secondary school is one step removed from university entrance and therefore less influential. Students entering from high school rated the clarity of educational and career goals more highly upon graduation than did students entering from college, and more highly in all three contexts than upper-year students. As well, students entering from college rated the criterion more highly at entry than did upper-year students. It would appear that acquired experience in the university setting attenuates the perceived

TABLE 6

Significant Effects of Gender, Program, and Cohort on Ratings of Criterion Importance

	Main and Interaction Effects			Multivariate post hoc Contrasts				
Criterion	Effect	*F*	*p*	Contrast H_0	Λ	*F*	*p*	Contrast Conclusion
General academic preparation	Gender	10.63	0.001	Male = Fem	0.969	4.05	0.007	F > M (Ent, Dur)
Sense of responsibility	Program	5.83	0.001	Arts = Sci	0.989	1.44	0.230	nsd
				Arts = Ed	0.956	5.79	0.001	Ed > A (all)
				Arts = Eng	0.976	3.02	0.030	nsd
				Sci = Ed	0.982	2.27	0.080	nsd
				Sci = Eng	0.989	1.42	0.237	nsd
				Ed = Eng	0.992	0.94	0.423	nsd
Clarity of student's educational and career goals	Program	5.03	0.002	Art = Sci	0.971	3.67	0.013	nsd
				Art = Ed	0.964	4.66	0.003	Ed > A (Ent, Grad)
				Art = Eng	0.962	4.89	0.002	Eng > A (Ent)
				Sci = Ed	0.988	1.58	0.193	nsd
				Sci = Eng	0.997	0.34	0.798	nsd
				Ed = Eng	0.995	0.68	0.563	nsd
	Cohort	5.12	0.006	1st HS = 1st col	0.971	3.68	0.012	1HS > 1Col (Grad)
				1st HS = upr	0.972	3.62	0.013	1HS > Upr (all)
				1st col = upr	0.969	3.95	0.009	1Col > Upr (Ent)
Secondary school preparation	Cohort	6.27	0.002	1st HS = 1st col	0.913	11.84	0.000	1HS > 1Col (Ent)
				1st HS = upr	0.979	2.66	0.048	nsd
				1st col = upr	0.933	8.90	0.000	Upr > 1Col (Ent, Dur)
Basic mathematical competency	Gen x Prog	4.26	0.006	$\text{Male}_{\text{art}} = \text{Fem}_{\text{art}}$	0.996	0.48	0.696	nsd
				$\text{Male}_{\text{sci}} = \text{Fem}_{\text{sci}}$	0.934	8.83	0.000	M-sci > F-sci (all)
				$\text{Male}_{\text{ed}} = \text{Fem}_{\text{ed}}$	0.997	0.33	0.804	nsd
				$\text{Male}_{\text{eng}} = \text{Fem}_{\text{eng}}$	0.996	0.52	0.666	nsd
				$\text{Art}_{\text{male}} = \text{Sci}_{\text{male}}$	0.942	7.67	0.000	S-m > A-m (all)
				$\text{Art}_{\text{male}} = \text{Ed}_{\text{male}}$	0.998	0.22	0.882	nsd
				$\text{Art}_{\text{male}} = \text{Eng}_{\text{male}}$	0.912	12.06	0.000	Eng-m > A-m (all)
				$\text{Sci}_{\text{male}} = \text{Ed}_{\text{male}}$	0.963	4.77	0.003	S-m > Ed-m (all)
				$\text{Sci}_{\text{male}} = \text{Eng}_{\text{male}}$	0.992	1.01	0.386	nsd
				$\text{Ed}_{\text{male}} = \text{Eng}_{\text{male}}$	0.950	6.52	0.000	Eng-m > Ed-m (all)

(Continued)

TABLE 6 (*Continued*)

Criterion	Main and Interaction Effects			Multivariate post hoc Contrasts				Contrast Conclusion
	Effect	F	p	Contrast H_0	Λ	F	p	
				$Art_{fem} = Sci_{fem}$	0.984	1.99	0.115	nsd
				$Art_{fem} = Ed_{fem}$	0.996	0.51	0.679	nsd
				$Art_{fem} = Eng_{fem}$	0.903	13.37	0.000	Eng-f > A-f (all)
				$Sci_{fem} = Ed_{fem}$	0.993	0.86	0.460	nsd
				$Sci_{fem} = Eng_{fem}$	0.942	7.66	0.000	Eng-f > S-f (all)
				$Ed_{fem} = Eng_{fem}$	0.919	10.94	0.000	Eng-f > Ed-f (all)

NOTES: Degrees of freedom for tests of effects and post hoc contrasts were: (a) *df*(gender) = 1, 376, (b) *df*(program) = 3, 376, (c) *df*(cohort) = 2, 376, (d) *df*(gender*program) = 3,376, and (e) *df*(post hoc) = 3, 374. Post hoc univariate results are based on tests of $F(1, 376)$. For each criterion, the critical level for statistical significance of main and interaction effects was set at $p \leq 0.006$, thereby maintaining an experiment-wise α of 0.09. Critical levels for post hoc contrasts were Bonferroni-adjusted, depending on the number of tests involved, in order to maintain an α of 0.05 for each set: (a) for $k = 3$ tests, $\alpha' = 0.018$, (b) for $k = 4$ tests, $\alpha' = 0.013$, (c) for $k = 6$ tests, $\alpha' = 0.009$.

A more detailed table is available from the first author upon request.

importance of goal clarity. For each cohort, however, the pattern was of increasing importance, suggesting that temporal effect outweighs cohort effect.

Basic mathematical competency presented the most complex situation: in addition to significant main effects for gender and program, there was a significant gender-by-program interaction. To investigate the consequences of the interaction for the main effects, post hoc contrasts were conducted in two phases. First, a series of contrasts examined gender differences within program groups. There was only one significant difference among the four programs; male science students rated the criterion more highly than did female science students in all three temporal contexts. Further investigation revealed that approximately half (46%) the male science students were from a physics course, and a similar proportion were from a psychology course. In comparison, 16% of the female science students were from the physics course and 84% were from the psychology course. Thus, the gender difference in ratings for science students might reflect differences between the physical and life sciences in the importance of mathematical competency. A repeated measures MANOVA of the perceived importance of basic mathematical competency between students in the physical and life sciences approached but did not reach statistical significance, $p = 0.073$.

A second series of contrasts examined program differences within gender groups. Among male students, those in engineering and science rated basic mathematical competency significantly higher in all three temporal contexts than did their counterparts in arts and education. Female engineering students rated the criterion significantly higher in all three temporal contexts than their counterparts in the other program areas. Based on these findings, we concluded that the gender by program interaction mitigated the main effects on importance ratings for basic mathematical competency.

Thus, among potential influences on the importance students attached to the criteria, program differences were found primarily for arts students, who attached less importance to clarity of educational and career goals and a sense of responsibility. Basic mathematical competency showed gender-by-program interactions. The remaining 22 criteria did not show program differences, which could be interpreted as consensus on their importance across programs.

Implications

Students' perceptions of the importance of criteria for university student quality confirmed those of a broader higher education sample of stakeholders, pointing to the general applicability of the criteria. The students considered all of the criteria to be important for assessing student quality, both in composite form and within each temporal context. Program of study, gender, and cohort affected the importance ratings of only a few criteria. Even in these cases, the criteria met or exceeded the threshold for importance as criteria of student quality.

The criteria sorted into meaningful principal components. Given the increasing attention to the effect of psychosocial factors such as involvement and motivation on student learning, it is noteworthy that the largest component contained both intellectual skills and psychosocial attributes. For example, this component included the ability to analyze, synthesize, and think critically; a sense of responsibility; the ability to interact with others; and self-confidence. We can deduce from these findings that students perceive themselves holistically, linking a wide range of personal characteristics. This stands in contrast to the serial and specific nature of their approach to academic preparation and performance.

Most striking from the results is that the majority of criteria (17/25) increased in importance from entry to university, during studies, to graduation. This leads first to the conclusion that when we speak of student quality we must acknowledge it to be a dynamic rather than a static phenomenon. It must be viewed in the context of a particular point in time during students' undergraduate education. The findings also suggest that students' perceptions of student quality are much more consistent with Astin's (1985) talent development view than with the more prevalent resource approach to excellence in postsecondary education. Students expect to gain in quality; although they perceive their entering characteristics (e.g., general academic preparedness) to be important, they place greater importance on what they will accomplish during their studies and upon graduation.

The results of this study suggest several avenues for further research. Differences in the importance of clarity of educational goals, particularly for arts students, signify a need for further exploration. Increasing attention is being paid to disciplinary differences in teaching and learning (for example, Hativa & Marincovich, 1995). More specifically differentiated program groups, discriminating between hard and soft, applied and pure, and life and non-life disciplines would increase explanatory power. Additional insight into program effects might be achieved, were students to make two sets of ratings for each criterion, one for university students generally, and a second for students in their particular program of study.

Improved understanding of the effects of temporal context requires use of more concrete or finite contexts; sampling for year of study rather than using the inclusive 'during studies' context for students overall would yield more specific ratings. Ideally, the same group of students would be surveyed each year throughout their academic studies (year of study as a within-subjects factor). Interview studies such as those conducted by Goulden and Griffin (1995) on the meaning of grades could be used to establish the range of student views and how the criteria are actually interpreted by them. Indicators developed in the Delphi study for each of the criteria (Donald & Denison, 1993) would aid in this analysis. Sampling students from varying types of institutions would reveal the extent to which students consider these criteria to apply to post-secondary students generally.

The results of this study suggest foremost that we must take into account the temporal context differences in the perceived importance of criteria. Greater consensus within individual institutions and in the broader higher education community about what is important and when must be achieved. Findings of differences across temporal contexts provide a stimulus for engaging in the necessary dialogue. What steps would we recommend be taken in response to students agreeing with the importance of the criteria, but clearly considering most criteria to be goals more than attributes of themselves at entry? Because cohort differences were minimal, we can conclude that students at entry are as aware of the importance of these criteria as are upper-level students. Indeed, upper-level students seemed to have lost some of the impetus for clear educational goals that incoming students displayed. Institutions may need to enunciate more clearly and support their goals for student learning in order to promote their development. The literature on student involvement and effort and their effects on student progress can serve as a guide (e.g., Astin, 1993, 1996; Pace, 1988,1990). The alignment and specification of institutional goals to ensure that students are committed to higher order learning and can find ways of supporting their commitment to their own development are challenges to be addressed.

References

Alexitch, L. R. (1997). Students' educational orientation and preferences for advising from university professors. *Journal of College Student Development, 38,* 333–343.

Astin, A. W. (1977). *Four critical years: Effects of college on beliefs, attitudes and knowledge.* San Francisco: Jossey-Bass.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25,* 297–308.

Astin, A. W. (1985). *Achieving educational excellence: A critical assessment of priorities and practices in higher education.* San Francisco: Jossey-Bass.

Astin, A. W. (1993). *What matters in college? "Four critical years" revisited.* San Francisco: Jossey-Bass.

Astin, A. W. (1996). "Involvement in learning" revisited: Lessons we have learned. *Journal of College Student Development, 37,* 123–134.

Baird, J. R. (1988). Quality: What should make higher education "higher"? *Higher Education Research and Development, 7,* 141–152.

Banta, T. W. (1989, March). *Campus-based assessment research: Examples of the genre.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Barnett, R. (1988). Entry and exit performance indicators for higher education: Some policy and research issues. *Assessment and Evaluation in Higher Education, 13,* 16–30.

Barzun, J. (1993). *The American university: How it runs, where it is going* (2nd ed.). Chicago: University of Chicago Press.

Baxter Magolda, M. B. (1990). The impact of the freshman year on epistemological development: Gender differences. *Review of Higher Education, 13,* 259–284.

Brennan, J., Fedrowitz, J., Huber, M., & Shah, T. (Eds.). (1999). *What kind of university? International perspectives on knowledge, participation and governance.* Buckingham UK: Open University Press with Society for Research into Higher Education.

Cantor, N., & Langston, C. A. (1989). Ups and downs of life tasks in a life transition. In L. A. Pervin (Ed.), *Goal concepts in personality and social psychology* (pp. 127–167). Hillsdale, NJ: Erlbaum.

Cantor, N., Norem, J., Langston, C., Zirkel, S., Fleeson, W., & Cook-Flanagan, C. (1991). Life tasks and daily experience. *Journal of Personality, 59,* 425–451.

Cashin, W. E., & Downey, R. G. (1995). Disciplinary differences in what is taught and in students' perceptions of what they learn and of how they are taught. In N. Hativa & M. Marincovich (Eds.), *Disciplinary differences in teaching and learning: Implications for practice.* New Directions for Teaching and Learning, No. 64 (pp. 81–92). San Francisco: Jossey-Bass.

Cave, M., & Hanney, S. (1992). Performance indicators. In B. R. Clark & G. R. Neave (Eds.) *The encyclopedia of higher education* (pp. 1411–1423). Oxford: Pergamon Press.

Chickering, A. W., & Reisser, L. (1993). *Education and identity* (2nd ed.). San Francisco: Jossey-Bass.

Constantinople, A. (1967). Perceived instrumentality of the college as a measure of attitudes toward college. *Journal of Personality and Social Psychology, 5,* 196–201.

Dey, E. L., Astin, A. W., & Korn, W. S. (1991). *The American freshman: Twenty-five year trends, 1966–1990.* Los Angeles: Higher Education Research Institute, Graduate School of Education, University of California.

Donald, J. G. (1990, June). *Criteria of quality and excellence in Canadian colleges and universities: Students.* Paper presented at the annual meeting of the Canadian Society for the Study of Higher Education, Vancouver, BC.

Donald, J. G. (1997). *Improving the environment for learning: Academic leaders talk about what works.* San Francisco: Jossey-Bass.

Donald, J. G. (1999). Motivation for higher order learning. In M. Theall (Ed.), *Motivation from within: Approaches for encouraging faculty and students to excel.*

New Directions for Teaching and Learning, No. 78 (pp. 27–35). San Francisco: Jossey-Bass.

Donald, J. G., & Denison, D. B. (1993, June). Criteria and indicators for quality and excellence in Canadian colleges and universities: The student domain. In *Criteria and indicators of quality and excellence in Canadian colleges and universities.* Symposium conducted at the annual meeting of the Canadian Society for the Study of Higher Education, Ottawa, ON.

Donaldson, E. L., & Dixon, E. A. (1995). Retaining women students in science involves more than course selection. *Canadian Journal of Higher Education, 25,* 29–51.

Ewell, P. T. (1991). To capture the ineffable: New forms of assessment in higher education. *Review of Research in Education, 17,* 75–125.

Feldman, K. A., & Newcomb, T. M. (1969). *The impact of college on students* (Vol. 1). San Francisco: Jossey-Bass.

Goulden, N. R., & Griffin, C. J. G. (1995). The meaning of grades based on faculty and student metaphors. *Communication Education, 44,* 110–125.

Hativa, N., & Marincovich, M. (1995). (Eds.) *Disciplinary differences in teaching and learning: Implications for practice.* New Directions for Teaching and Learning, No. 64. San Francisco: Jossey-Bass.

Herndon, J. B., Kaiser, J., & Creamer, D. G. (1996). Student preferences for advising style in community college environments. *Journal of College Student Development, 37,* 637–648.

King, P. M., & Baxter Magolda, M. B. (1996). A developmental perspective on learning. *Journal of College Student Development, 37,* 163–173.

Kuh, G. D., Schuh, J. H., Whitt, E. J., & Associates (1991). *Involving colleges: Successful approaches to fostering student learning and development outside the classroom.* San Francisco: Jossey-Bass.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications.* Reading, MA: Addison-Wesley.

Martin, E., Ramsden, P., & Bowden, J. (1989, November). *Students' conceptions of adaptation to higher education.* Paper presented at the annual meeting of the Australian Association for Research in Education, Adelaide, Australia.

McGill University. (1996). *Report on enrollment statistics, 1995–1996.* Montreal: Author.

Mentkowski, M., & Loacker, G. (1985). Assessing and validating the outcomes of college. In P. T. Ewell (Ed.), *Assessing educational outcomes.* New Directions for Institutional Research, No. 47 (pp. 47–64). San Francisco: Jossey-Bass.

Meyer, J. H. F., Dunne, T. T., & Richardson, J. T. E. (1994). A gender comparison of contextualised study behaviour in higher education. *Higher Education, 27,* 469–485.

Nadeau, G. G., Donald, J. G., & Konrad, A. (1992, April). *Criteria and indicators of quality and excellence in Canadian colleges and universities.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Pace, C. R. (1988). *Measuring the quality of college student experiences.* Los Angeles: UCLA Center for the Study of Evaluation.

Pace, C. R. (1990). *The undergraduates: A report of their activities and progress in college in the 1980s.* Los Angeles: UCLA Center for the Study of Evaluation.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Pintrich, P. (1995). (Ed.) *Understanding self-regulated learning.* New Directions for Teaching and Learning, No. 63. San Francisco: Jossey Bass.

Pintrich, P. R., & Garcia, T. (1994). Self-regulated learning in college students: Knowledge, strategies, and motivation. In P. R. Pintrich, D. R. Brown, & C. E. Weinstein (Eds.), *Student motivation, cognition, and learning: Essays in honor of Wilbert J. McKeachie* (pp. 113–133). Hillsdale, NJ: Erlbaum.

Schroeder, C. C., & Hurst, J. C. (1996). Designing learning environments that integrate curricular and cocurricular experiences. *Journal of College Student Development, 37,* 174–181.

Terenzini, P. T. (1989). Assessment with open eyes: Pitfalls in studying student outcomes. *Journal of Higher Education, 60,* 644–664.

Trice, A. G., & Dey, E. L. (1997). Trends in faculty teaching goals: A longitudinal study of change. *Journal of College Student Development, 38,* 527–534.

Volet, S. E. (1997). Cognitive and affective variables in academic learning: The significance of direction and effort in students' goals. *Learning and Instruction, 7,* 235–254.

Volet, S. E., & Lawrence, J. A. (1990). Goals in the adaptive learning of university students. In H. Mandl, E. De Corte, N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction: European research in an international context: Vol. 2.1. Social and cognitive aspects of learning and instruction* (pp. 497–516). Oxford: Pergamon.

Willis, D. (1993). Academic involvement at university. *Higher Education, 25,* 133–150.

Book Reviews

University Teaching: International Perspectives,
edited by James J. F. Forest. Garland Publishing,
1998. 432+ pp. $87.00

LESLIE ADAMS WIMSATT, University of Michigan

If recent international conference programs are any indication, the quality of teaching in higher education is attracting the attention of policymakers and researchers around the world. In many countries, concerns about teaching arise in the context of increasing and more heterogeneous student enrollments, decreasing levels of funding, and a resultant push for "budgetary efficiency" and productivity. In some cases, a sharpened focus on teaching is also driven by technological innovation and marketplace competitiveness. Researchers and administrators are left to wonder how these trends and the resulting changes in teaching policy are affecting the culture and practice of teaching in different countries. Hence, an examination of recent international developments in university teaching is both timely and necessary.

University Teaching is a collection of eighteen independent essays and case studies that explore several dimensions of university teaching, including the impact of cultural and social environments. The book, edited by James Forest, is divided into four topic-specific sections: (1) issues of instruction, (2) perspectives on student learning and assessment, (3) training and development of university teachers, and (4) issues of policy, structure, and organization. Chapters are authored by an international group of faculty and administrators with significant experience in the areas of teaching, learning, and assessment at the postsecondary level. Although it includes far too much information to highlight in a brief review, this book is a good resource for readers interested in examining how issues of university teaching transcend national boundaries.

W. Alan Wright opens "Issues of Instruction" by identifying international policies and practices most likely to influence teaching quality (e.g., value of teaching in personnel decisions, role of academic leadership, impact of mentoring and instructional development programs). He then uses survey responses from instructional developers and academics to create a design for teaching improvement. Using comparative research to investigate how students in different Western and non-Western cultures conceptualize "good teaching," David Watkins further develops the theme of this section. Watkins shows how an inter-institutional project in Hong Kong effectively employs research-based indicators of college quality to enhance student learning and finds that Western concepts about quality of instruction also hold in non-Western cultures. James Forest concludes this discussion by synthesizing a large volume of theory and research. His treatment of factors that influence the environment for teaching and learning is extensive, including such measures as diversity, communication, attitudes and behavior, instructional methodology and evaluation, and technological advancement.

"Perspectives on Student Learning and Assessment" targets student outcomes in relation to faculty teaching. Noel Entwistle begins with an extensive literature review, which synthesizes European and Australasian theory and research and culminates with the introduction of a conceptual overview of the teaching-learning process. His observations about how student characteristics, faculty teaching methods, and institution-based policies and practices are linked to student learning outcomes is a useful counterpoise to France Mugler and Roger Landbeck's case study of distance learning at a regional university in the South Pacific. Entwistle's theoretical approach to strategic planning is in sharp contrast to Mugler and Landbeck's case-specific plan for the improvement of distance education outcomes by redesigning study materials, assessment instruments and academic support services. Two other chapters in this section present innovative approaches to teaching and assessment. Kari Smith discusses portfolio assessment in the context of Israeli teacher training courses and describes how portfolios can be used as both a learning tool and as a method of student assessment. Ted and Patricia Panitz review literature identifying the benefits and pitfalls of collaborative learning in the university classroom and suggest ways to overcome faculty resistance. Exploring issues of authority, self-disclosure, and cultural relevance linked to student learning in Micronesia, Kyle Smith, Seyda Tòrk Smith, and Iain Twaddle urge university faculty to recognize how levels of collectivism implicit in course materials may be inconsistent with student approaches to learning. The authors recommend that course designs take into account students' culturally based approaches to learning as well as attitudes toward faculty-student interactions.

The third section of the book supplements theoretical discussion with several practical models for the "Training and Development of University Teachers." Graham Gibbs notes that practices for the preparation of university teachers are rapidly developing in scale and sophistication in Europe and Australasia due to increasing professionalization and the support of formal policies. In contrast, he reports that teacher preparation efforts in the United States are considerably more limited in scope. Offering specific programming information, Michael Herrick explains the process of conducting a faculty development program in Botswana, including how an assessment instrument structured around eight components of effective teaching is used to evaluate teaching performance. In the final chapter, John Dwyer calls for a renewed commitment toward the development and recognition of good teaching, applauding one university in Canada for developing a university practicum specifically designed to train graduate students as teachers.

The final section of the book focuses on issues of policy, structure, and organization. Margaret Robertson notes how benchmarks are increasingly used in Australian universities for measuring teaching performance and allocating competitive teaching funds. Concluding that better measures are needed to capture the current realities of students' learning environments, her ideas contribute to current debate about the impact of performance evaluation, "vocationalization," technological innovation, and changing administrative cultures on teaching. Terry Hyland discusses work-based learning and professional higher education in Britain in the context of a broad review of extant literature. He worries that workplace learning converts university education into job training and inhibits institutional teamwork, strategic collaboration, and community service. Finding that educational opportunity can easily become a struggle between quantitative and qualitative objectives, Tronie Rifkin questions the efficacy of quantitative assessment instruments to evaluate academic program participants. Rifkin cites the case of a three-year baccalaureate degree program in Denmark and concludes that such assessment practices can lead to an overemphasis on productivity and efficiency. Howard Fergus offers an overview of the organizational history and policy implications of distance education in the West Indies and lobbies for increased academic support, revised teaching materials, and expanded teacher training. In a theoretically grounded approach to the integration of writing across the curriculum, Sheila Vance and Glenda Crosling highlight the importance of placing the design of instruction and evaluation within disciplinary and national contexts. They cite two case studies to illustrate how innovative writing programs have been designed and successfully implemented in Australia. Yung Che Kim's chapter on the effect of national contexts on teaching in South Korea directly addresses the role of student activism. He asserts that college teaching, instructional development, and evaluation need to be culturally and socially relevant to students and faculty in order to be effective. Finally, Marcela Mollis and Daniel Feldman discuss contested academic appointments at the University of Buenos Aires and then consider how institutional policies can affect the culture of teaching within individual academic departments.

University Teaching makes an important contribution to the comparative study of higher education because the extant literature has tended to focus on academic systems and policy. Academic administrators and policy makers will appreciate the pragmatic suggestions for improving institutional teaching polices and practices and researchers will find beneficial some extensive and thorough reviews of relevant theory and research. Because each chapter makes an individual contribution to the literature, the book can be used in whole or in part. It requires the reader to tie together emergent themes, key findings, and conclusions rather than to rely on introductory and summary chapters. Although the book could be more consistent in the reporting of research results, I found it particularly informative and useful in documenting how specific programs and practices are being used to improve teaching in different countries.

The authors wisely suggest that longitudinal examination of teaching policy and practice is needed to distinguish short-term gains from more enduring international reforms. Persistent questions remain regarding the impact of student diversity, faculty characteristics, institutional and disciplinary contexts, personal beliefs, and individual activities on teaching practices. Also, David Watkins' essay brings to mind questions about the appropriateness of current indicators of college quality. For example, how can more specific measures of college quality be used to evaluate institutional commitment to curricular reform and innovation?

University Teaching complements and supplements other useful works from the comparative higher education literature, including Altbach's (1998) *Comparative Higher Education,* which examines the complex relationships among academic systems in countries with varying levels of economic and technological development; and Clark's (1995) *Places of Inquiry,* which conceptualizes the contemporary university from a research vantage point to examine how national trends either strengthen or weaken the integration of research and teaching. These three books provide an overview of some fundamental issues affecting teaching on an international level. Taken together they also reveal the potential for organizational change as a function of teaching improvement.

References

Altbach, P. G. (1998). *Comparative higher education: Knowledge, the university, and development.* Greenwich, CT: Ablex.

Clark, B. R. (1995). *Places of inquiry: Research and advanced education in modern universities.* Berkeley: University of California Press.

Condemning Students to Debt: College Loans and Public Policy,
edited by Richard Fossey and Mark Bateman. New York: Teachers
College Press, 1998. 224 pp. $38.00

SCOTT R. SWEETLAND, The Ohio State University

The student loan has been conceptualized as a financial aid since its federal inception in 1965. Its purpose has been to fill the gap that often exists between the total cost of higher education and non-obligatory financial aids that are provided to many students. The problem is that contemporary higher education policies have begun to substitute the student loan for an increasing proportion of non-obligatory financial aids to students, creating new generations of debtors who may be least able to afford an increasing debt burden. That is the major theme of this book.

Condemning Students to Debt includes ten chapters contributed by eleven authors. It is well organized and easy to read, yet each chapter is written with scholarly style to support further review and inquiry. Among the chapters, one can find the history behind student loans and student financial aid; reasons for the shift in public policy away from student financial aid and toward student loans; potential impacts of this shift on the higher education establishment; economic implications and contemporary fiscal decisions faced by students and their families; and the perplexity of possible solutions that can be effected by public policy and institutional administration. Of particular interest to me is the survey and explanation of economic contexts contained in Chapter 2. Here, the justification and need for governmental involvement in higher education investment considers monetary and non-monetary benefits that accrue to society as well as to individuals and their families when postsecondary education programs are financially supported by governments. The case is made that the student loan can be used as an important public policy tool that mutually improves the societal values or objectives of equity and efficiency.

The theme of equity or equal educational opportunity transcends from across all the chapters to support the student loan program but also to question its efficacy in degree and implementation. Not only has the total amount on-loan increased over recent decades but many of the contributing authors assert that an increased reliance on the student loan program has essentially replaced other more equity-oriented financial grants-in-aid. Moreover, this effect can deter economically disadvantaged students from pursuing higher education degrees while making the same pursuits more affordable to students who are, in the economic sense, relatively less dependent upon student loans and financial grants-in-aid. Unfortunately, when the student loan repayment horizon is considered in separate chapters on default and bankruptcy, it is readily apparent that economically disadvantaged students who borrow money to finance their postsecondary studies are most at risk throughout the entire investment-repayment cycle. That is, for some incoming freshmen, factors of economic disadvantage can contribute to academic hardship and difficulty in the labor market and later exacerbate financial problems, if the repayment of student loan obligations becomes difficult or impossible. Given the evidence present in these chapters, the replacement of equity-oriented financial grants-in-aid with student loan obligations seriously limits the probability that important equity aims will be achieved.

My only disappointment in this book is the overall lack of attention paid to technical aspects of the student loan mechanism. The contributing authors cite volumes of earliest as well as current literature and data that address relevant problems, policies, implications, and possibilities, but the book, after all, is about loans. The value of the book would increase with a chapter dedicated to the theoretical framework and practical features of loans to include such terms and concepts as amortization, inflation, origination fees, basis points, nominal and real interest rates, guarantee, subsidy, hedging, matching, and credit analysis. Although the authors mention many of these in describing their analyses, other important lines of inquiry would be opened through greater attention to the more technical aspects. For example, my student loan portfolio includes debt that was incurred during the 1980s at 7% interest and at 8.25% interest during the 1990s. Although the increased rate of interest was modest, corresponding home mortgage interest rates were about 15% and 7%, suggesting that interest rates on student loans were highly subsidized during the 1980s but that more recent rates of interest on student loans actually include a premium relative to home mortgage interest rates. If my experience is equivalent to the national average experience, then the financial burden associated with student loans has increased over time in addition to the burdens posed by total and proportional increases in student loan amounts that have been observed and reported by the authors.

There can be no question that student loans have become an increasingly popular means of supporting higher education expenditures. But at what point will this trajectory end? Many contributions in this book synthesize a call for clarity in the definition of public policy objectives that differentiate student loan obligations from financial grants in student aid. Otherwise, the trend toward student loans, in substitution for non-obligatory financial aids, may lead to new generations of student debtors who are unable to fulfill their financial obligations. This book provides a strong foundation for further research into this potentiality and other important public policy issues that directly affect higher education and the national economy.

Instructions to Contributors

Manuscripts should be mailed in triplicate to Leonard L. Baird, Editor, *The Journal of Higher Education,* Ohio State University Press, 1070 Carmack Road, Columbus, OH 43210-1002.

Style. *The Journal of Higher Education* has adopted as its official guide the *Publication Manual of the American Psychological Association,* fourth edition, and all manuscripts should be brought into conformity with this guide before they are submitted. Papers should be typed, double-spaced, on white 8½ × 11-inch paper, with wide margins. Upon acceptance, contributors will be asked to supply a computer file on an IBM-compatible standard disk in WordPerfect. Authors who use a different system should prepare an ASCII text file. An abstract of fifty words or less, summarizing the main points of the article, should accompany the manuscript. Because the journal's readers represent a variety of professional interests, it is recommended that any statistical material be presented as briefly and simply as possible. Although each paper submitted should deal with the methodology employed in addressing the subject in sufficient detail to place the data within the proper methodological setting, the editors of the journal are not primarily interested in papers setting forth practices of research methods (an acquaintance with fundamental procedures of scholarly analysis being assumed on the part of the reader) except for those papers that develop innovative methodological approaches.

Illustrations submitted with the final draft must be of professional quality, and executed on white paper or vellum, in black ink, with clear, medium weight, black lines and figures. Typewritten lettering should not appear in illustrations. Figures should be provided **at size**—no larger than 4½ × 7 inches (full page) and preferably no larger that 4½ × 3½ inches (half page) and **printed camera-ready at a minimum of 600 dpi.** They should be numbered consecutively, and the number and author's name should be penciled lightly on the back of each. All illustrations must have captions, which should not appear on the artwork but should be typed, double-spaced, on a sheet at the end of the manuscript. If there is any potential for doubt, the word *top* should be written on the back of the illustration.

Authors should employ a Reference List format to list bibliographic data. Endnotes should be reserved for supplementary comment and typed on a separate page at the end of the manuscript.

Manuscript Length. Manuscripts averaging 25–30 pages of double-spaced typescript are preferred, but the editors will consider longer papers on topics requiring a fuller treatment.

Review Process. Those unsolicited manuscripts that are refereed are reviewed blind. Authors are thus requested to submit their name, professional position, and institution on a removable cover sheet. They should also mask any items of self-reference where they appear. Authors must not submit the manuscript of any article that is still under consideration by another publisher.

Editorial Reaction. Papers will not be returned to authors if they fail to meet by a wide margin the basic criteria for selection. Otherwise, authors may expect to receive some notification within three months. If an article is accepted, it will usually appear in print within twelve months after acceptance. If an article that has been subjected to a full review is rejected, the opinions of the referees will be transmitted to the author.

Criteria for Selection. Papers are evaluated on the following points: Form: writing style and readability, logical development, appropriate length, appropriateness of author's stated objectives to treatment as defined below. Content: significance to *JHE* readers. Additional criteria are based upon the following manuscript orientations: as a research paper, as a technical paper, as a professional practice paper, as a literature review, and as a policy paper. It should be emphasized that the editors respond most favorably to manuscripts that evidence both a freshness of vision and a vitality that may be informed by, but certainly go beyond, methodological qualities, and that are in congruence with our publishing goals and directions. The most effective approach in learning about our interests is to read previous issues of the journal. We expect that authors, the journal, and the field will develop through the publication process.

Patrick T. Terenzini
Alberto F. Cabrera
Carol L. Colbeck
Stefani A. Bjorklund
John M. Parente

Racial and Ethnic Diversity in the Classroom

Does It Promote Student Learning?

Since passage of the Civil Rights Act of 1964 and the Higher Education Act of 1965, America's colleges and universities have struggled to increase the racial and ethnic diversity of their students and faculty members, and "affirmative action" has become the policy-of-choice to achieve that heterogeneity. These policies, however, are now at the center of an intense national debate. The current legal foundation for affirmative action policies rests on the 1978 *Regents of the University of California v. Bakke* case, in which Justice William Powell argued that race could be considered among the factors on which admissions decisions were based. More recently, however, the U.S. Court of Appeals for the Fifth Circuit, in the 1996 *Hopwood v. State of Texas* case, found Powell's argument wanting. Court decisions turning affirmative action policies aside have been accompanied by state referenda, legislation, and related actions banning or sharply reducing race-sensitive admissions or hiring in California, Florida, Louisiana, Maine, Massachusetts, Michigan, Mississippi, New Hampshire, Rhode Island, and Puerto Rico (Healy, 1998a, 1998b, 1999).

An earlier version of this article was presented at the meeting of the Association for the Study of Higher Education, San Antonio, Texas, November 1999. The study was supported in part by a grant from the National Science Foundation (Grant No. 634066D) to the Engineering Coalition of Schools for Excellence in Education and Leadership (ECSEL). The opinions expressed here do not necessarily reflect the opinions or policies of the National Science Foundation or the ECSEL Coalition, and no official endorsement should be inferred.

Patrick T. Terenzini is professor and senior scientist; Alberto F. Cabrera is associate professor and senior research associate; Carol L. Colbeck is assistant professor and research associate; Stefani A. Bjorklund is a graduate research assistant; and John M. Parente is a graduate research assistant. Center for the Study of Higher Education, The Pennsylvania State University.

The Journal of Higher Education, Vol. 72, No. 5 (September/October 2001)
Copyright © 2001 by The Ohio State University

In response, educators and others have advanced educational arguments supporting affirmative action, claiming that a diverse student body is more educationally effective than a more homogeneous one. Harvard University President Neil Rudenstine claims that the "fundamental rationale for student diversity in higher education [is] its educational value" (Rudenstine, 1999, p. 1). Lee Bollinger, Rudenstine's counterpart at the University of Michigan, has asserted, "A classroom that does not have a significant representation from members of different races produces an impoverished discussion" (Schmidt, 1998, p. A32). These two presidents are not alone in their beliefs. A statement published by the Association of American Universities and endorsed by the presidents of 62 research universities stated: "We speak first and foremost as educators. We believe that our students benefit significantly from education that takes place within a diverse setting" ("On the Importance of diversity in University Admissions," *The New York Times*, April 24, 1997, p. A27).

Studies of the impact of diversity on student educational outcomes tend to approach the ways students encounter "diversity" in any of three ways. A small group of studies treat students' contacts with "diversity" largely as a function of the numerical or proportional racial/ethnic or gender mix of students on a campus (e.g., Chang, 1996, 1999a; Kanter, 1977; Sax, 1996). Gurin (1999) and Hurtado, Milem, Clayton-Pedersen, and Allen (1999) refer to this numerical or proportional "mix" of students as "structural diversity." Whether such diversity is a *sufficient* condition to promote student educational outcomes, however, is far from clear.

A second, considerably larger set of studies take some modicum of structural diversity as a given and operationalize students' encounters with diversity using the frequency or nature of their reported interactions with peers who are racially/ethnically different from themselves. In these studies, which might be labeled "*in situ* diversity studies," encountering diversity is viewed as part of the normal processes and functioning of campus life or of a campus's racial/ethnic and gender climate (e.g., Antonio, 1998; Astin, 1993; Cabrera, Nora, Terenzini, Pascarella, & Hagedorn, 1999; Davis, 1994; Gurin, 1999; Pascarella, Edison, Nora, Hagedorn, & Terenzini, 1996; Whitt, Edison, Pascarella, Nora, & Terenzini, 1999).

A third set of studies examines institutionally structured and purposeful programmatic efforts to help students engage racial/ethnic and/or gender "diversity" in the form of both ideas and people. This category includes studies of the influences of coursework and the curriculum (e.g., Astin, 1993; Chang, 1999b; Cohen, 1994; Cohen, Bianchini, Cossey, Holthuis, Morphew, & Whitcomb, 1997; Hurtado, 1999; MacPhee, Kreutzer, & Fritz, 1994; Palmer, 1999), and participation in racial or multicultural awareness workshops (e.g., Antony, 1993; Astin, 1993; Springer, Palmer, Terenzini, Pascarella, & Nora, 1996; Vilalpando, 1994), as well as various other forms of institutional programming intended to enhance the diversity of a campus or the educational consequences of engaging "diversity" in one form or another (see Musil, Garcia, Moses, & Smith, 1995; Rendon & Hope, 1996; Sedlacek, 1995). Appel, Cartwright, Smith, and Wolf (1996), Smith (1989), and Hurtado et al. (1999) provide useful reviews of this literature.

These various approaches have been used to examine the effects of diversity on a broad array of student educational outcomes. The evidence is almost uniformly consistent in indicating that students in a racial/ethnically or gender-diverse community, or engaged in a diversity-related activity, reap a wide array of positive educational benefits. "Diversity" in its various forms has been linked to such outcomes as higher minority student retention (e.g., Bowen & Bok, 1998; Chang, 1996, 1999a), greater cognitive development (e.g., Adams & Zhou-McGovern, 1994; Cohen, 1994; Cohen, et al., 1997; Hurtado, 1999; MacPhee et al., 1994; Sax, 1996), and positive gains on a wide-range of measures of interpersonal and psychosocial developmental changes, including increased openness to diversity and challenge (Pascarella, et al., 1996), greater racial/cultural knowledge and understanding and commitment to social justice (Antonio, 1998; Astin, 1993; Chang, 1999b; Milem, 1994; Palmer, 1999; Springer, et al., 1996), more positive academic and social self-concepts (Astin, 1993; Chang, 1996; Sax, 1996), more complex civic-related attitudes and values, and greater involvement in civic and community-service behaviors (Astin, 1993; Milem, 1994; Hurtado, 1999). (Chang [1998] and Milem [1999] provide excellent reviews of this literature.)

As noted above, however, only a relative handful of studies (e.g., Chang, 1996, 1999a; Sax, 1996) have specifically examined whether *the racial/ethnic or gender composition* of the students on a campus, in an academic major, or in a classroom (i.e., structural diversity) has the educational benefits claimed by Rudenstine, Bollinger, and others. Sax found that the proportion of women in an academic major field had no impact on students' cognitive or affective development. Chang's analyses reveal a good bit of the complexity of the relation between structural diversity, student interactions and experiences, and educational outcomes. He found that a campus's racial heterogeneity had an effect on learning outcomes through its influence on students' diversity-related experiences, specifically, socializing with peers from different racial/ethnic backgrounds and discussing racial/ethnic issues. Whether the degree of racial diversity of a campus or classroom has a *direct* effect on learning outcomes, however, remains an open question. The scarcity of information on the educational benefits of the structural diversity on a campus or in its classrooms is regrettable because it is the sort of evidence the courts appear to be requiring if they are to support race-sensitive admissions policies.

In addition to the shortage of information on the role of structural diversity, most studies examine diversity's influence on various dimensions of students' psychosocial development, including (but not limited to) racial/ethnic attitudes and values, academic and social self-concepts, civic behaviors, and racial/ethnic awareness and knowledge. Far fewer studies (e.g., Cohen, 1994; Cohen, et al., 1997; MacPhee et al., 1994; Slavin, 1995) explore the influence of diversity in the classroom or in other small groups on students' development of academic or intellectual knowledge and skills.

This study attempted to contribute to the knowledge base by exploring the influence of structural diversity in the classroom on students' development of academic and intellectual skills. The study put to an empirical test Bollinger's claim that racially/ethnically homogeneous classrooms produce "an impoverished" educational experience (Schmidt, 1998, p. A32). The study was designed to evaluate whether and to what extent (if any) the racial/ethnic diversity of the students in a classroom is related to student learning, specifically, to gains in students' problem-solving skills and their abilities to work in groups. In addition, this study sought to extend Chang's (1996, 1999a) work indicating that structural diversity was associated with more frequent, diversity-related experiences which, in turn, were related to educational outcomes. This study examines both the direct effect of classroom diversity on academic/intellectual outcomes and whether any effects of classroom diversity may be moderated by the extent to which active and collaborative instructional approaches are used in the course.

Methods

Conceptual Underpinnings

In this study, we assume that the development of students' course-related skills are shaped by students' precourse characteristics, the instructional practices encountered in the classroom, and the racial/ethnic diversity of the classroom. Students' precourse characteristics are assumed to be temporally prior to both classroom diversity and instructional methods in their effects on learning outcomes. Our primary focus is on the influence of varying levels of classroom diversity on students' learning outcome *above and beyond the effects of other variables that may* *also influence learning* (e.g., students' precourse characteristics and the pedagogical methods adopted by instructors).

Sample and Data Collection

This study was part of an evaluation of the National Science Foundation-funded Engineering Coalition of Schools for Excellence in Education and Leadership (ECSEL). ECSEL comprises seven colleges of engineering: City College of New York, Howard University, the Massachusetts Institute of Technology, Morgan State University, Pennsylvania State University, the University of Maryland, and the University of Washington. Among other goals, ECSEL seeks to promote the use of design groups, or engineering teams, throughout the undergraduate curriculum in helping students learn to solve unstructured engineering problems. The original data collection was intended to evaluate the extent to which the active and collaborative learning activities inherent in group-based engineering design promoted student learning when compared with more traditional approaches to teaching (e.g., lecture and discussion).

The base sample consists of 1,258 engineering students enrolled at all 7 ECSEL institutions who completed the Classroom Activities and Outcomes Survey (described below). Participating courses and students were not randomly selected. The local ECSEL evaluator on each campus was asked to identify as many "ECSEL" courses (in which design was being taught using active and collaborative learning techniques) as feasible, as well as (for comparative purposes) several "non-ECSEL" courses with educational goals similar to those of the ECSEL courses. In the non-ECSEL courses, traditional lecture and discussion techniques were the primary mode of instruction.

Survey forms were administered in 49 classrooms. Of these, 29 were ECSEL classes, and 20 were non-ECSEL classes. Of the 1,258 students, 936 (74%) were enrolled in an ECSEL course while 322 (26%) were in non-ECSEL courses. Because of the nonrandom nature of the data collection, 46% of the students were enrolled at the University of Maryland, 21% at the University of Washington, and 13% at The Pennsylvania State University. The remaining 20% were distributed approximately evenly across the City College of New York, Howard University, Morgan State University, and MIT. The analyses reported here are based on the responses of 680 white students (58% of the sample) and 488 students of color. In the overall sample, 180 respondents (15.4%) were African Americans, 234 (20.0%) were Asian Americans, 64 (5.5%) were Latino/as, and 10 (0.9%) were Native Americans. Students were approximately evenly distributed across class years, with 57% in lower-division courses and 43% in upper-division courses. No significant differences in this distribution were identified between ECSEL and non-ECSEL course students. While the total database for this study contained the original 1,258 students, the *n*'s for the several analyses described below varied between 962 and 1,194 because of missing data. Because of the relatively large number of cases with missing data on some variables, it was decided to drop those cases from analyses rather than use mean replacement.

Instrument and Variables

The data for this study come from the Classroom Activities and Outcomes Survey, a pencil-and-paper, multiple-choice questionnaire completed at the end of a course. The instrument has three sections. The first gathers information on students' personal and academic backgrounds and demographic characteristics. The second section asks about the characteristics and activities of the course in which the students were enrolled when completing the questionnaire. The final section asks students about the extent to which they believe they have made progress in various learning and skill development areas *as a result of taking that particular course*. (A copy of the Classroom Activities and Outcomes Survey is available from the first author at <ptt2@psu.edu>.)

Control variables. Background characteristics controlled in this study included gender (coded: 1 = male, 0 = female), race/ethnicity (coded: 1 = nonminority, 0 = minority; group *n*'s did not permit disaggregation of race/ethnicity into more discrete categories), and high school academic achievement (combined SAT scores).

Independent variables. The second section of The Classroom Activities and Outcomes Survey asks students to report how often during the course they or their instructor engaged in each of 26 classroom activities. Respondents use a 4-point scale, where 1 = never, 2 = occasionally, 3 = often, and 4 = very often/almost always. The items comprising this section were drawn from the research literature on effective instructional practices and activities.

A principal components factor analysis of these 26 items (with varimax rotation) produced 5 factors. This solution, accounting for 62.2% of the variance in the correlation matrix, is shown in Table 1. Three of the five factors related to specific instructional practices. Collaborative Learning consists of 7 practices that reflect the interdependence among students required by working in groups. The Instructor Interaction and Feedback factor included 5 practices that fostered frequent, supportive communication between faculty and students. The 3-item Clarity and Organization factor reflects instructors' use of clear explanations and an integrated course structure. The fourth and fifth factors contained 2 and 4 items, respectively, reflecting students' perceptions of fairness in the treatment of minorities and women in the classroom by the faculty member (the Faculty Climate scale) and by other students (the Peer Climate scale). As can be seen at the bottom of the table, the internal consistency reliabilities (Cronbach's alpha) for these scales were generally high, ranging from 0.77 to 0.89. The classroom climate measures were excluded from the set of independent variables to provide a more precise estimation of the effects of classroom diversity on learning unconfounded by students' perceptions of racial or gender dynamics in the classroom, which might, in themselves, affect learning. In affirmative action cases, moreover, the courts' interest has been specifically in the educational contributions (if any) of the racial/ethnic composition of the learning setting.

TABLE 1

Factor Structures for Classroom Practice Items

	Factor Loadings				
Items	Collaborative Learning	Instructor Interaction & Feedback	Clarity & Organization	Faculty Climate	Peer Climate
Discuss ideas with classmates	0.822				
Work cooperatively with students	0.739				
Opportunities to work in groups	0.753				
Get feedback from classmates	0.753				
Students teach & learn from one another	0.679				
Interact with classmates outside of class	0.650				
Require participation in class	0.589				
Interact with instructor as part of the course		0.780			
Interact with instructor outside of class		0.741			
Instructor gives *detailed* feedback		0.713			
Instructor gives *frequent* feedback		0.689			
Guided student learning versus lecturing		0.578			
Assignments/activities clearly explained			0.767		
Assignments/presentations clearly related			0.722		
Instructor makes clear expectations for activities			0.677		
Instructor treats minorities the same as whites				0.913	
Instructor treats women the same as men				0.901	
In groups, some males treat women differently					0.876
Some male students treat women differently					0.869
Some white students treat minorities differently					0.865
In groups, some whites treat minorities differently					0.809
Internal Consistency Reliability (Alpha)	0.88	0.83	0.77	0.86	0.89

Classroom diversity, the independent variable of principal interest in this study, was operationalized using a "diversity index" created by dividing the number of students who reported their racial/ethnic identity to be non-white by the total number of students in the class. Because two ECSEL institutions are Historically Black Universities, the diversity index was calculated so that classrooms with a diversity "mix" approaching 50% were considered the most diverse. Classrooms with a percentage of students of color lower than 50% (or, in the case of HBCU classrooms, greater than 50%) were considered to be less diverse. For example, a classroom in which all students were white *or* all were students of color was considered to have no diversity.

A preliminary examination indicated that the distribution of the diversity index was curvilinear (i.e., as classroom diversity increased, the nature of the effect on reported learning gains changed). In order to examine the nature and effects of this nonlinear relation more easily, the diversity index was used to develop five categories of "classroom diversity." Table 2 shows the five categories, the ranges of the classroom diversity levels within each category, and the number and percentage of students who were in classes falling within each category. For example, about 40% of the students were in courses characterized as "medium" diversity classrooms. This category contains students in predominantly white courses in which 33–38% of the total enrollment were students of color *as well as* students in predominantly minority-student courses in which 33–38% of all the students were white. The categories were formed using natural breaks in the multi-modal frequency distribution. With the exception of the "medium" diversity category, which is the largest group, respondents were distributed relatively evenly across the five categories.

Dependent variables. The third part of the Classroom Activities and Outcomes Survey asks students to report the progress they believe they have made in 27 areas *as a result of the course for which they were completing the survey form.* Progress is reported on a 1-to-4 scale, where

TABLE 2

Classroom Diversity Categories, Intervals, and Number and Percentage of Students in Each Group

Categories	Classroom Diversity Intervals	Students *n*	Students %
No diversity	0% or 100%	142	11.3%
Low diversity	6–19%	184	14.7
Medium–low diversity	22–30	185	14.7
Medium diversity	33–38	500	39.7
High diversity	40–50	247	19.6
TOTALS		1,258	100.0%

1 = none, 2 = slight, 3 = moderate, and 4 = a great deal. These items were drawn primarily (but not exclusively) from a series of Delphi studies by Jones and her colleagues (Jones, 1994; Jones, et al., 1994) intended to develop consensus among faculty members, research specialists, academic administrators, and employers on definitions and components of "critical thinking" and "problem solving."

A principal components factor analysis (with varimax rotation) of the 27 skill development items yielded three factors: Problem-Solving Skills (12 items), Group Functioning Skills (7 items), and Occupational Awareness (4 items). This three-factor solution explained 64.6% of the total item variance and produced scales with internal consistency reliabilities ranging from 0.81 to 0.93. The composition of these factors is given in Table 3. Because of the interest in this study in students' skill development, the Occupational Awareness scale was excluded from further analyses.

TABLE 3

Factor Structures for Learning Outcome Items

	Factor Loadings		
Items	Group Skills	Problem-Solving Skills	Occupational Awareness
Developing ways to resolve conflict & reach agreement	0.779		
Being aware of feelings of members in group	0.841		
Listening to the ideas of others with open mind	0.829		
Working on collaborative projects as member of a team	0.815		
Organizing information to aid comprehension	0.679		
Asking probing questions that clarify facts, concepts	0.606		
Developing alternatives that combine best from previous work	0.618		
Ability to do design		0.578	
Solve an unstructured problem		0.697	
Identify knowledge, resources, & people to solve problem		0.666	
Evaluate arguments & evidence of competing alternatives		0.675	
Apply an abstract concept or idea to a real problem		0.735	
Divide problems into manageable components		0.744	
Clearly describe a problem orally		0.679	
Clearly describe a problem in writing		0.667	
Develop several methods to solve unstructured problem		0.732	
Identify tasks needed to solve an unstructured problem		0.752	
Visualize what the product of a design project would look like		0.584	
Weigh the pros/cons of possible solutions to a problem		0.623	
Understanding what engineers do			0.754
Understanding language of design			0.721
Understanding engineering has a nontechnical side			0.710
Understanding of the process of design			0.703
Internal Consistency Reliability (Alpha)	0.926	0.943	0.813

For both the Classroom Activities and Skill Development Outcome factors, scales were created by summing students' responses on a factor's component items and then dividing by the number of items the factor contains.

Analytical Methods

Ordinary least-squares multiple regression analyses were used in a series of hierarchical analyses. First, to determine whether the diversity of the classrooms had *any* association with learning outcomes, each of the two dependent variables (self-reported gains in problem-solving and group skills) was regressed on four of the five levels of classroom diversity (students in courses with no diversity constituted the reference group). Second, reported gains in problem-solving and group skills were again regressed on classroom diversity after controlling for students' race/ethnicity, gender, and academic ability. Third, each learning outcome was regressed hierarchically on: (1) students' race/ethnicity, gender, and ability, (2) the three scales reflecting the instructional methods used in the classroom (collaborative learning, instructor interaction and feedback, and course clarity and organization), and (3) four levels of classroom diversity.

Finally, the influence of classroom diversity may well be contextual, that is, conditional (or dependent) on the degree to which students interact with one another in course-related activities. For example, interpersonal contacts among students in low diversity courses may well have a different effect on learning than similar contacts in medium- or high-diversity classrooms. To evaluate the extent to which classroom diversity's effects may vary depending on the instructional methods used, a set of four cross-product interaction terms was created by cross-multiplying each of the four levels of classroom diversity (low through high) by students' scores on the Collaborative Learning scale. These interaction terms were then entered as a set into an OLS regression (one for each dependent variable) after students' precourse characteristics, instructional methods, and the four diversity levels had been entered as main effects variables.

Results

Table 4 reports the means and standard deviations for each group's reported gains in problem-solving and group skill development, although the relations among (and magnitudes of the differences between) the group means for both outcomes are more easily seen in Figures 1 and 2.

The basic patterns of the relations between classroom diversity and reported learning are the same for both problem-solving and group skill development. As can be seen in both figures, and with students in "no diversity" classrooms as the reference group, the reported gains drop to their lowest among students in "low diversity" classrooms, although those drops are not statistically significant (as indicated by the "n.s." between the data points). The trend line then climbs through the mean for "low-medium" diversity classrooms, peaking among students in "medium" diversity classes, only to fall again for students in "high diversity" courses. The magnitudes of the differences between the various diversity levels are similar and relatively small. Only the differences between medium-level course means and (with one exception) all other group means are statistically significant (based on pair-wise Scheffé *post hoc* comparisons). (The exception is that for the problem-solving outcome, students in medium diversity classrooms report gains at approximately the same level as do students in classrooms with no diversity.)

The results of each of the three phases of the analyses are reported in Table 5. Consistent with the analyses underlying Figures 1 and 2, the first-phase regressions (with levels of classroom diversity as the only predictors) indicate that classroom diversity is, indeed, related to students' self-reported development of both their problem-solving and group skills. While statistically significant ($p < 0.001$), however, the overall relation in both analyses is small (adjusted R^2s of 0.02 and 0.05 for the problem-solving and group skill models, respectively). The beta weights indicate that (relative to the reference group: students in classes with no diversity) low classroom diversity is negatively related to students' development of both problem-solving and group skills at statistically significant levels. In the regression on group skills, moreover, both low and medium-low levels of classroom diversity are significantly and negatively related to reported gains. It is worth noting that in the group

TABLE 4

Means and Standard Deviations on Learning Outcomes by Level of Classroom Diversity ($n = 1{,}199$)

	Problem-Solving Skills		Group Skills	
Classroom Diversity Level	Mean	*SD*	Mean	*SD*
No diversity	2.82	0.75	2.81	0.88
Low diversity	2.68	0.68	2.57	0.83
Medium–low diversity	2.77	0.63	2.71	0.81
Medium diversity	2.96	0.64	3.05	0.64
High diversity	2.77	0.67	2.72	0.83
TOTALS	2.84	0.67	2.84	0.78

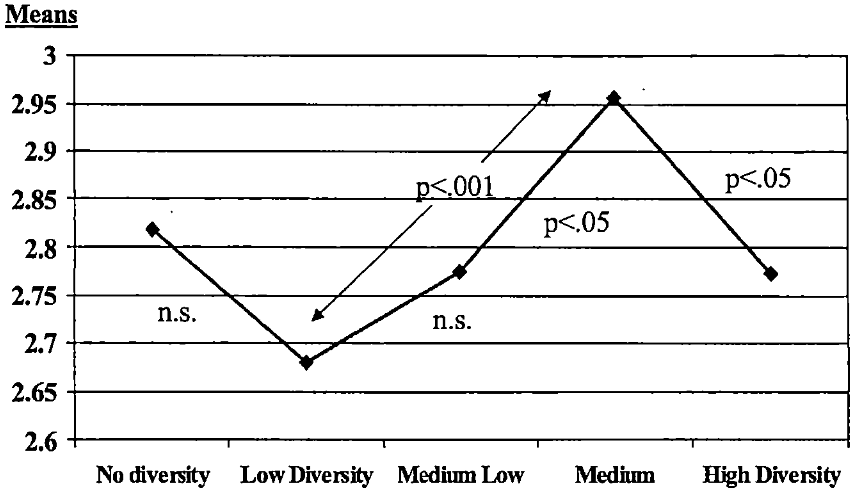


FIG. 1. Differences in Group Means for Gains in Problem-Solving Skills

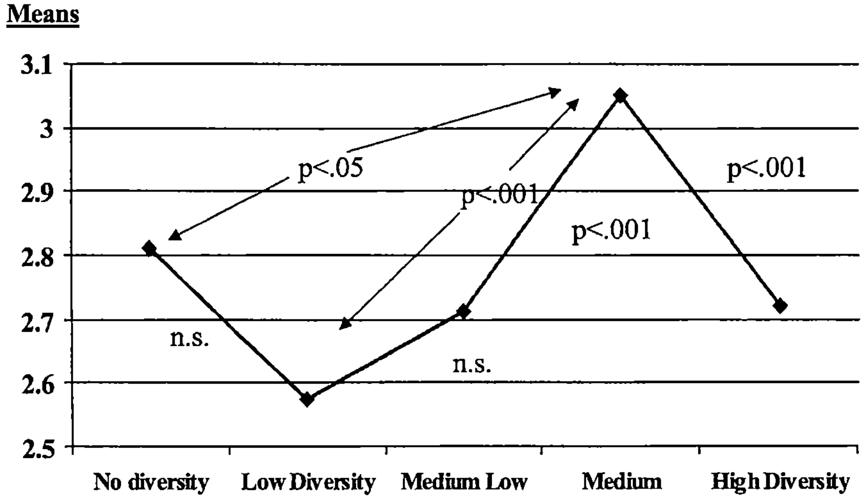


FIG. 2. Differences in Group Means for Gains in Group Skills

skills model, the beta weights for both medium and high levels of diversity (0.07 and –0.09, respectively) approach traditional levels of statistical significance ($p < 0.17$ and $p < 0.07$, respectively), with medium diversity positively related to reported group skill learning gains, while a high level of diversity is negatively related to students' reported gains.

TABLE 5

Results of the Three Phases of Regression Analyses

	Problem-Solving Skills		Group Skills	
Independent Variables	Adj. R^2	Betas	Adj. R^2	Betas
Classroom Diversity Only	0.02***		0.05***	
Low diversity		–0.12**		–0.16***
Medium-low diversity		–0.05		–0.09*
Medium diversity		0.04		0.07
High diversity		–0.04		–0.09
Precourse Characteristics and Classroom Diversity	0.02***		0.05***	
Gender		0.03		–0.05
Race/ethnicity		–0.04		–0.00
Ability (SATs)		–0.03		–0.02
Low diversity		–0.13**		–0.15***
Medium-low diversity		–0.06		–0.08
Medium diversity		0.03		0.09
High diversity		–0.04		–0.08
Precourse Characteristics, Classroom Activities, and Classroom Diversity	0.34***		0.34***	
Gender		0.04		–0.02
Race/ethnicity		–0.01		0.03
Ability		–0.07*		–0.08**
Collaborative learning		0.28***		0.41***
Instructor interaction		0.32***		0.20***
Clarity & organization		0.13***		0.08**
Low diversity		0.00		–0.01
Medium-low diversity		–0.01		–0.01
Medium diversity		0.06		0.09
High diversity		0.04		0.00

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

In the second phase analyses (reported in the middle portion of Table 4), despite the addition of controls for students' race/ethnicity, gender, and academic ability, the association between classroom diversity and both learning outcomes persists relatively unchanged. The adjusted R^2s remain low and, indeed, are identical (within rounding error) to those in the first-phase models. Again, low levels of classroom diversity (relative to no diversity at all) were negatively related to gains in both problem-solving and group skills at statistically significant levels. The pattern of the signs of the beta weights also remained unchanged, suggesting that medium-low and high levels of classroom diversity may be negatively related to reported learning gains, while medium levels of diversity appear to have a positive effect on learning. None of these weights, however, reached the traditional standard of statistical significance. In the group skills model, however, the weights for medium-low, medium, and high levels of classroom diversity did approach the traditional criterion of statistical significance ($p < 0.08$, 0.11, and 0.13, respectively). In sum, gender, race/ethnicity, and academic ability appear to have no statistically significant effect on the learning reported by students in this study, while the evidence continues to suggest that classroom diversity may be a factor (possibly both positive and negative).

In the third phase of the analyses, with students' precourse characteristics and the three classroom activity scales included as control variables, the adjusted R^2 values climb substantially to 0.34 for both problem-solving and group skills. The sharp rise in the R^2 was predictable. Both the research literature and common sense would lead one to expect the addition of three scales reflecting what goes on in a classroom to be powerful predictors of how much students think they have learned. Indeed, one might reasonably expect these close-to-the-action predictors to eliminate what the earlier regressions have suggested is the relatively small contribution of classroom diversity to learning gains. Such expectations were largely—but not completely—borne out. As can be seen in the bottom portion of Table 4, none of the beta weights for the various levels of classroom diversity reached statistical significance. Classroom diversity, as a factor in student learning, however, did not disappear entirely. In the model predicting reported gains in problem-solving skills, the diversity index, when treated as a *continuous* variable (rather than being treated as dichotomous categories), produced a beta weight of 0.045. Although small relative to the weights of the three classroom activity scales, the diversity index weight remained statistically significant ($p < 0.05$). Similarly, in the group skills model, the beta weight for medium levels of diversity (0.09) was statistically significant at $p < 0.067$, only narrowly failing to meet the conventional $p < 0.05$ standard. Thus, these findings indicate that what happens in a classroom (e.g., the degree to which students engage in active and collaborative learning activities, their interactions with instructors and peers, and the level of clarity and organization in the classroom) are clearly more powerful influences on students' reported learning gains than is the level of the classroom's structural diversity. Nonetheless, classroom diversity, *despite* the presence of these more proximal and powerful influences, continued to have a measurable influence on student learning (albeit a small and statistically marginal one by conventional standards).

Finally, entry of the set of four diversity-by-collaborative learning scale interaction terms produced no appreciable increase in the value of R^2 for either the problem-solving or group skills models. Thus, the data in this study suggest that the effects of the level of classroom diversity on students' reported skill-development gains are general rather than conditional (or dependent) on the use of collaborative learning approaches in the course.

Limitations

This study has several limitations. First, although the sample is multi-institutional and contains a broad range of engineering schools, the seven institutions that participated in the study were not randomly selected. Thus, to an unknown degree, these institutions may not be representative of the national mix of engineering schools or, indeed, of all four-year universities. Similarly, the classes and students invited to participate in the study were not randomly selected. Although local evaluators were urged to sample ECSEL and non-ECSEL courses from across their institution's college of engineering's class levels, the resulting samples may not be representative of the course or student populations (engineering or otherwise) on each campus. Moreover, the number of classes and students participating vary widely across the participating institutions. Thus, generalizations to other institutions' engineering classes and students must be made cautiously. With regard to sampling, however, the study has two distinct assets when compared to most studies of classroom effects on student learning: its multi-institutional design and its relatively large sample of both courses and students.

Second, the influences of gender and academic ability are probably underestimated in this study due to the relative homogeneity of engineering students on these campuses with respect to these variables. By and large, the participants in this study were male (73%) and academically very able (mean combined SATs of 1,241).

Third, while problem-solving and group skills are basic educational outcomes of most engineering (and general education) programs, they are certainly not the only dimensions along which future engineers (or students in general) develop academically and intellectually during their undergraduate programs. Moreover, alternative conceptualizations and operationalizations of "problem-solving" and "group" skills have been advanced, and the results of this study might have been somewhat different had other conceptualizations and/or measures of each skill been used, or if other, entirely different learning outcomes had been the foci of this study.

Fourth, the measurements of skill development in this study are based on students' self-reports rather than on more objective measures of student learning (e.g., standardized tests). Recent research suggests, however, that self-report measures of learning can be used to appraise gains in cognitive skills. Pike (1995) found self-reported measures of educational gains to be as valid as objective measures to the extent that the self-report measures reflect the content of the learning outcome under consideration. As noted earlier, the items reflecting the learning outcomes studied in this research came primarily (albeit not exclusively) from a national study of the beliefs of faculty members, researchers, administrators, and employers about what component abilities make up those skills (Jones, 1994; Jones, et al., 1994). Similarly, Anaya (1999), after examining a representative sample of students who took the Graduate Record Examinations in 1989, concluded that self-reported measures of gains in cognitive skills are valid proxies of cognitive skills as measured by the verbal and math components of the GRE. Moreover, while standardized measures have some advantages over self-reports, they also come with limitations of their own for classroom use, including availability, length, cost, and relevance to specific courses. The self-report instrument used in this study was designed specifically to gather course-level information and to be easy and inexpensive to use. One must, nonetheless, acknowledge the trade-offs being made.

Fifth, the study's design and database are cross-sectional. The concept of learning "gains" or skill "development" implies change over time. Moreover, the impact of course- (or campus-) diversity may also vary over time. A longitudinal design would provide a more rigorous test of whether classroom diversity is related to learning. It is worth noting in this regard, however, that the relations identified between classroom diversity and reported learning gains persisted in the presence of controls for selected precourse student characteristics (gender, race/ethnicity, and high-school academic achievement) and remained marginally significant even in the presence of psychometrically sound measures of classroom activities designed to promote learning. These latter measures, as one might expect, were clearly more powerful forces for student learning, but they failed to completely eradicate evidence that classroom diversity may also be involved.

Sixth, for reasons explained earlier relating to the apparent curvilinear effect of classroom diversity on reported learning gains, this study operationalized classroom diversity as four, dummy-coded levels, rather than as a single, continuous variable. Subsequent use of these dummy-coded variables in statistical interaction terms to examine whether classroom diversity might have a different effect depending on the degree of interpersonal contact among students in the course provides a relatively low-power test of the possible, conditional effects of diversity on learning. Future studies should examine more rigorously the possibility of such "contextual" effects. The structural diversity in a classroom (and elsewhere) may, indeed, have a general effect (i.e., one that is about the same, regardless of classroom activities), rather than "conditional" or "contextual" (i.e., one in which the magnitude of the effect varies depending on the setting), but for theoretical, practical, and policy reasons, that relation should be validated.

Seventh, this study examined the influence of different levels of classroom diversity only in relation to the effect of no classroom diversity at all. Some levels of diversity, independent of the kinds of pedagogies adopted, may be more or less, positively or negatively, related to learning gains. This study shed no light on these questions, and future research on the matter is strongly encouraged.

Finally, students develop their problem-solving and group skills over time and at varying rates. This study is limited by the fact that changes in these skill areas were examined after only one course. The *cumulative* changes in these areas that can be attributed to the racial/ethnic diversity in these and subsequent courses throughout students' academic programs, as well as in their out-of-class encounters with racially and ethnically diverse individuals, may be more extensive than what is reported here. Indeed, one might also hypothesize that the overall institutional climate for diversity is a more powerful force for learning than is the level of diversity in individual classrooms. Chang's (1996, 1999a) work supports this proposition, but it offers no insight into the relative influence of campus- vs. classroom-level diversity. Because this study was unable to control for campus-level diversity climate, the hypothesis that the campus climate is the dominant force remains a plausible alternative to the interpretation of the findings in this study. It is worth noting, however, that the phrasing of the survey items consistently reminded students that they were being asked to describe the activities going on in a specific course and to report learning gains associated specifically with that course. Moreover, Cabrera and Nora (1994) report findings consistent across racial/ethnic groups that students' sense of institutional alienation is shaped more powerfully by their in-class experiences than by their perceptions of the general campus climate.

Conclusions and Implications

Since the passage of the Civil Rights Act of 1964 and the Higher Education Act of 1965, America's colleges and universities have struggled to provide equal access to applicants of all races and ethnicities. Affirmative Action, based on racially and ethnically sensitive admissions decision making, has been the policy of choice in trying to achieve equality of access and racially and ethnically diverse student bodies.

Widely adopted as it has been, however, affirmative action has become increasingly controversial. Reliance on race-sensitive admissions received the support of the U.S. Supreme Court in the 1978 *University of California v. Bakke* decision, when Justice William Powell, writing for the majority, argued that race could be one of the factors on which admissions decisions were based. The *Bakke* decision came under fire, however, in the 1996 *Hopwood v. Texas* case when the U.S. Court of Appeals for the Fifth Circuit rejected arguments supporting the University of Texas' use of race-sensitive admissions to its law school. That ruling was subsequently extended to all admissions activities in Texas' public higher education systems, and it has shaped referenda or legislative actions in a number of other states nationwide.

In response, representatives of colleges and universities have argued that affirmative action is necessary to maintain racially and ethnically diverse student bodies and that the practice is defensible on educational, if not legal, grounds. Diverse student bodies and classrooms, the argument goes, are more educationally effective than are less- or non-diverse ones. Lee Bollinger, president of the University of Michigan, for example, has asserted, "A classroom that does not have a significant representation from members of different races produces an impoverished discussion" (Schmidt, 1998, p. A32).

A growing body of research has lent support to this argument, although the evidence is far from conclusive. A significant segment of this literature focuses on the effects of a campus's racial/ethnic climate on students' racial/ethnic attitudes or learning. These studies are generally consistent in finding that a "warmer" climate is related to students' willingness to socialize and discuss racial issues and to greater tolerance and appreciation for diverse populations. A second segment of the diversity research has examined the effectiveness of specific, programmatic initiatives (e.g., cultural awareness workshops and diversity course requirements) intended to promote greater tolerance and understanding among racially and ethnically diverse students. Like the campus climate research, this body of evidence generally supports the effectiveness of such programmatic interventions.

Few studies, however, specifically examine whether the racial/ethnic composition of a campus or classroom—the so-called "structural diversity" of these settings—has a measurable impact on student learning. This study explored precisely that question with respect to the racial/ethnic composition of individual classrooms, as well as whether the effects of structural diversity might be mediated by the kinds of instructional methods in use in the classroom. The findings of this study hardly constitute a ringing endorsement of Bollinger's claim that "a classroom that does not have a significant representation from members of different races produces an impoverished discussion" (Schmidt, 1998, p. A32). Portions of the evidence do, however, support claims about the educational benefits of racially or ethnically diverse classrooms. Level of classroom diversity was related at small—but statistically significant—levels to students' reported gains in both their problem-solving and their group skills. Moreover, those relations persisted even in the presence of controls for students' race/ethnicity, gender, and academic ability. In the most rigorous tests applied in this study, both students' precourse characteristics (including ability) and three scales reflecting the instructional practices in use in the course were controlled, the level of classroom diversity *continued* to show a positive, if small, statistically marginal relation to reported learning gains. In a regression on students' reported gains in problem-solving skills, a continuous measure of classroom diversity had a small, but statistically significant, positive effect (beta = 0.045, $p < 0.05$). In a similar regression on students' reported gains in their group skills, being in a "medium diversity" classroom was positively related to reported gains net of other student characteristics, instructional methods, and other levels of classroom diversity. This effect failed, but only narrowly ($p < 0.07$), to meet the conventional standard for statistical significance. These findings indicate quite clearly that what happens in a course is far-and-away a more powerful predictor of learning outcomes than is the level of classroom diversity. Nonetheless, the persistence of diversity's influence *despite* the presence of more powerful predictors is, we believe, substantively noteworthy and relevant to the policy question this study seeks to illuminate.

The evidence also suggests that the relation between the racial/ethnic composition of a classroom and students' learning gains may not be a simple, linear one. The evidence quite consistently indicates that "medium" levels of classroom diversity (i.e., approximately in the 30–40% range) are positively, if not always significantly, related to students' reports of learning gains. At the most rudimentary level of analysis, however, the data also suggest the possibility that low or high levels of classroom diversity *may* be negatively related to learning gains. Analyses examining the effects of only classroom diversity level, or of diversity level while controlling for students' race/ethnicity, gender, and ability, produced some marginal evidence of statistically significant but negative effects among students in classrooms with low or high levels of diversity (compared to no diversity at all). These negative relations, however, were not supported when measures of the instructional practices in use in these classrooms entered the analyses. Entry of a set of interaction terms (reflecting whether the effects of various levels of diversity varied depending on the extent to which collaborative learning activities were used in the classroom) produced no appreciable increase in the value of the R^2 for either model. This finding suggests that any effects of structural diversity appear to be general and not conditional on the instructional methods in use in the classroom. That conclusion, however, warrants further validation. Similarly, future research should examine more precisely than was possible here the levels at which classroom diversity becomes a salient positive or negative force in shaping students' learning.

At best, the findings in this study suggest a small, if statistically significant, link between the level of racial/ethnic diversity in a classroom and students' reports of increases in their problem-solving and group skills. The relatively consistent and positive salience of medium levels of classroom diversity is the most supportive evidence for arguments that classroom diversity has positive, educational effects on student learning. That evidence, however, is far from conclusive.

The findings of this study are all the more suggestive when one considers that the relation between diversity and student learning is at least modestly detectable *in individual classrooms.* One might reasonably speculate that, if the beneficial effects of racial/ethnic diversity are apparent in *individual* classrooms, then those effects may well be substantially magnified in the aggregate, when accumulated across the courses students take and across their out-of-class experiences in racially/ethnically diverse settings.

Should subsequent studies of the effects of the racial/ethnic composition of classrooms and other campus settings bear out the relations suggested in this research, then much of the current cloudiness in the legal and policy worlds concerning the educational effectiveness of diverse settings may be clarified. Arguments for affirmative action and race-sensitive admissions that assert the educational effectiveness of such policies will rest on substantially firmer empirical ground, and campus, state, and federal policies permitting or encouraging race/ethnicity-sensitive admissions will also rest on firmer empirical ground

References

Adams, M., & Zhou-McGovern, Y. (1994, April). *The sociomoral development of undergraduates in a "social diversity" course.* Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Anaya, G. (1999). College impact on student learning: Comparing the use of self-reported gains, standardized test scores, and college grades. *Research in Higher Education, 40,* 499–527.

Antonio, A. L. (1998). *The impact of friendship groups in a multicultural university.* Unpublished doctoral dissertation, University of California, Los Angeles.

Antony, J. (1993, November). *Can we all get along? How college impacts students' sense of promoting racial understanding.* Paper presented at the meeting of the Association for the Study of Higher Education, Pittsburgh, PA.

Appel, M., Cartwright, D., Smith, D. G., & Wolf, L. E. (1996). *The impact of diversity on students: A preliminary review of the research literature.* Washington, DC: Association of American Colleges and Universities.

Astin, A. W. (1993). *What matters in college?* Four critical years *revisited.* San Francisco: Jossey-Bass.

Bowen, W. G., & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions.* Princeton, NJ: Princeton University Press.

Cabrera, A. F., & Nora, A. (1994). College students' perceptions of prejudice and discrimination and their feelings of alienation: A construct validation approach. *Review of Education, Pedagogy, Cultural Studies, 16,* 387–409.

Cabrera, A. F., Nora, A., Terenzini, P. T., Pascarella, E. T., & Hagedorn, L. S. (1999). Campus racial climate and the adjustment of students to college: A comparison between white students and African-American students. *Journal of Higher Education, 70,* 134–160.

Chang, M. J. (1996). *Racial diversity in higher education: Does a racially mixed student population affect educational outcomes?* Unpublished doctoral dissertation, University of California, Los Angeles.

Chang, M. J. (1998). *An examination of conceptual and empirical linkages between diversity initiatives and student learning in higher education.* Unpublished manuscript prepared for the American Council on Education's Symposium and Working Research Meeting on Diversity and Affirmative Action, Arlington, VA, January 1999.

Chang, M. J. (1999a). Does racial diversity matter? The educational impact of a racially diverse undergraduate population. *Journal of College Student Development, 40,* 377–395.

Chang, M. J. (1999b, November). *The impact of an undergraduate diversity requirement on students' level of racial prejudice.* Paper presented at the meeting of the Association for the Study of Higher Education, San Antonio, TX.

Cohen, E. G. (1994). *Designing group work: Strategies for heterogeneous classrooms* (2nd ed.). New York: Teachers College Press.

Cohen, E. G., Bianchini, J. A., Cossey, R., Holthuis, N. C., Morphew, C. C., & Whitcomb, J. A. (1997). What did students learn? 1982–1994). In E. G. Cohen & R. A. Lotan (Eds.), *Working for equity in heterogeneous classrooms.* New York: Teachers College Press.

Davis, J. E. (1994). College in black and white: Campus environment and academic achievement of African American males. *Journal of Negro Education. 63,* 620–633.

Gurin, P. (1999). Empirical results from the analyses conducted for this litigation. From *Expert report of Patricia Gurin: Gratz, et al., v. Bollinger, et aL, No. 97–75321* (E.D. Mich.), *Grutter, et al. v. Bollinger, et al.,* No. 97–75928 (E.D. Mich.). http://www.umich.edu/~urel/admissions/legal/expert/empir.html

Healy, P. (1998a, November). Advocates of more money for colleges win key gubernatorial races. *Chronicle of Higher Education,* pp. A29–A30.

Healy, P. (1998b, December). U.S. Appeals Court ruling may imperil university defenses of affirmative action. *Chronicle of Higher Education,* pp. A30–A31.

Healy, P. (1999, March). University of Massachusetts limits racial preferences, despite vow to increase minority enrollment. *Chronicle of Higher Education,* pp. A30–A31.

Hurtado, S. (1999). Linking diversity and educational purpose: How the diversity of the faculty and student body impacts the classroom environment and student development. In G. Orfield (Ed.), *Diversity challenged: Legal crisis and new evidence.* Cambridge, MA: Harvard Publishing Group.

Hurtado, S., Milem, J., Clayton-Pedersen, A., & Allen, W. (1999). *Enacting diverse learning environments: Improving the climate for racial/ethnic diversity in higher education.* (ASHE-ERIC Higher Education Report, Vol. 26, No. 8). Washington, DC: The George Washington University, Graduate School of Education and Human Development.

Jones, E. A. (1994). *Goals inventories.* University Park, PA: The Pennsylvania State University, Center for the Study of Higher Education, National Center on Postsecondary Teaching, Learning, and Assessment.

Jones, E. A., with Hoffman, S., Moore, L. M., Ratcliff, G., Tibbetts, B, & Click, B. A. (1994). *Essential skills in writing, speech and listening, and critical thinking: Perspectives of faculty, employers, and policy makers.* University Park, PA: The Pennsylvania State University, Center for the Study of Higher Education, National Center on Postsecondary Teaching, Learning, and Assessment.

Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology, 82,* 965–990.

MacPhee, D., Kreutzer, J. C., & Fritz, J. J. (1994). Infusing a diversity perspective into human development courses. *Child Development, 65,* 699–715.

Milem, J. F. (1994). College, students, and racial understanding. *Thought & Action, 9,* 51–92.

Milem, J. F. (1999). The educational benefits of diversity: Evidence from multiple sectors. In M. Chang, D. Witt, J. Jones, & K. Hakuta (Eds.), *Compelling interest: Examining the evidence on racial dynamics in higher education.* Report of the AERA Panel on Racial Dynamics in Colleges and Universities. Palo Alto, CA: Stanford University Center for Comparative Studies in Race and Ethnicity. (Prepublication copy, available at http://www.stanford.edu/~hakuta/RaceInHigherEducation.html.)

Musil, C., Garcia, M., Moses, Y., & Smith, D. (1995). *Diversity in higher education: A work in progress.* Washington, DC: Association of American Colleges and Universities.

Palmer, E. A. (1999). *An analysis of a diversity requirement: The effects of course characteristics on students' racial attitudes, gender attitudes, and self-perceived learning.* Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.

Pascarella, E. T., Edison, M., Nora, A., Hagedorn, L. S., & Terenzini, P. T. (1996). Influences on students' openness to diversity and challenge in the first year of college. *Journal of Higher Education, 67,* 174–195.

Pike, G. R. (1995). The relationship between self-reports of college experiences and achievement test scores. *Research in Higher Education, 36,* 1–22.

Rendon, L. I., & Hope, R. (1996). *Educating a new majority.* San Francisco: Jossey-Bass.

Rudenstine, N. L. (1999). *Why a diverse student body is so important.* http://www.inform.umd.edu/diversity_web/Profiles/divdbase/harvard /ilsc.html

Sax, L. J. (1996). The dynamics of "tokenism:" How college students are affected by the proportion of women in their major. *Research in Higher Education, 37,* 389–425.

Schmidt, P. (1998, October). University of Michigan prepared to defend admissions policy in court. *Chronicle of Higher Education,* p. A32.

Sedlacek, W. (1995). *Improving racial and ethnic diversity and campus climate at four-year independent Midwest colleges: An evaluation report of the Lilly endowment Grant Program.* College Park, MD: University of Maryland—College Park.

Slavin, R. (1995). Cooperative learning groups and intergroup relations. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural education.* New York: MacMillan.

Smith, D. G. (1989). *The challenge of diversity: Involvement or alienation in the academy?* (ASHE-ERIC Higher Education Reports, No. 5). Washington, DC: George Washington University Press.

Springer, L., Palmer, B., Terenzini, P. T., Pascarella, E. T., & Nora, A. (1996). Attitudes toward campus diversity: Participation in a racial or cultural awareness workshop. *Review of Higher Education, 20,* 53–68.

Vilalpando, O. (1994, November). *Comparing effects of multiculturalism and diversity on minority and white students' satisfaction with college.* Paper presented at the meeting of the Association for the Study of Higher Education, Tucson, AZ. (ERIC Document Reproduction Service No. ED 375 721)

Whitt, E. J., Edison, M., Pascarella, E. T., Nora, A., & Terenzini, P. T. (1999). Women's perceptions of a "chilly" climate and cognitive outcomes in college: Additional evidence. *Journal of College Student Development, 40,* 163–177.

ȷE Douglas T. Shapiro

Modeling Supply and Demand for Arts and Sciences Faculty

What Ten Years of Data Tell Us About the Labor Market Projections of Bowen and Sosa

Introduction: Labor Market Models, Data, and Critiques

In 1989 William Bowen and Julie Sosa published *Prospects for Faculty in the Arts & Sciences*, attempting to project the balance of faculty supply and demand over a 25-year span, from 1987 through 2012. The book was the last of a number of similar reports and studies (e.g., Bowen & Schuster, 1986; Lozier & Dooris, 1987; McGuire & Price, 1989), all of which predicted, from the depths of a recession in the academic labor markets, that a rapid turnaround was in sight. Although the supply of new PhDs had far exceeded the demand for new faculty throughout the eighties, these works were intended as wake-up calls to the fact that the large numbers of faculty who had been hired during higher education's expansion in the sixties would soon retire, creating a swelling demand for new faculty that would be difficult to fill at the then current rate of PhD production. The intent was to spur action on the part of institutions and policymakers to increase the flow of students into graduate programs and rebuild the supply of faculty.

All of these works relied on "fixed-coefficient" models for calculating future supply and demand. That is, they projected current trends mathematically, without attempting to account for feedback loops or adjustment mechanisms of the marketplace. Yet, there is a range of different levels of sophistication. Whereas much of the previous work had been based upon relatively simple assumptions, Bowen and Sosa took a much

An earlier version of this article was presented at the annual meeting of the Association for the Study of Higher Education, San Antonio, November 1999.

Douglas T Shapiro is a doctoral candidate at the Center for the Study of Higher and Postsecondary Education at the University of Michigan.

The Journal of Higher Education, Vol. 72, No. 5 (September/October 2001)
Copyright © 2001 by The Ohio State University

more detailed approach, making *Prospects for Faculty* extremely rich in hard data. Indeed, its authors had access to levels of detail in their data that had been almost unimaginable just a few years earlier. Bowen and Schuster (1986), for example, were unapologetic in their work's aggregation of the entire national academic labor market, explaining that "the necessary data . . . [to break the totals down into components such as geographic regions, types of institutions, or academic disciplines] . . . are not available and we do not have the resources to gather them" (p. 195). In their defense, Bowen and Schuster were primarily concerned with the quality of faculty, not quantity. Bowen and Sosa, on the other hand, focusing almost entirely on quantity, were able to obtain finer data and disaggregate their figures, both for faculty and for graduate degrees, by individual academic fields and by type of institution. Although their model required them to make many assumptions about the future values of various factors and coefficients, the majority of these assumptions are carefully supported by detailed analyses of data trends.

Out of a universe of, at one count, 96 distinct scenarios that might be played out over the quarter century of their vision, Bowen and Sosa eventually narrowed their projections down to four main models that they considered most likely to occur. With only slight variation, each of those models projected significant shortages of PhDs in the academic labor market arising as early as 1992 and by 1997 at the latest, reversing an oversupply of PhDs during the years leading up to that point. Unlike their contemporaries, however, Bowen and Sosa did not attribute the projected shortages to any sudden wave of retirements, but rather to the combined effects of slight recoveries in postsecondary enrollments (after a drop in the early nineties), stagnant production of PhDs (with actual declines in the percentage of new PhDs who choose to pursue academic careers) and most importantly, a large, steady outflow of faculty from the combined life processes of career changes, retirements, and deaths.

Almost immediately after its publication, some researchers began to criticize the accuracy and reliability of the projections made in *Prospects for Faculty* (Blum, 1991; Gill, 1992), even as others rushed to proclaim the arrival of the first faculty shortages years ahead of when Bowen and Sosa had predicted they would begin (El-Khawas, 1990, 1991). It is a measure of the authority and credibility of Bowen and Sosa's work, however, that even those most critical of their methods generally balked at rejecting their conclusions. Gill (1992), for example, pointed to the unreliability of the data upon which Bowen and Sosa's projections were based and cautioned that some of their key assumptions regarding both enrollments and sources of faculty supply might be invalid, yet they went on to promote five urgent initiatives that institutional policy makers should pursue "to prepare effectively for a possible faculty shortage" (p 37).

Unfortunately, the next few years brought no attempts to test Bowen and Sosa's assumptions against the real marketplace, nor to revise or improve upon their effort. Instead, the focus of researchers turned to the economic recession and the impact of its attendant institutional retrenchments. According to Jack Schuster, coauthor of *American Professors: A National Resource Imperiled* (1986), "It's pretty simple what happened. We were blindsided by an abrupt economic downturn in 1990 and 1991. The ripple effect went everywhere and colleges and universities began to . . . curtail faculty hiring" (quoted in Brodie, 1995). Analysts chose to look no further than state budget cuts for explanations of the failure of *Prospects for Faculty's* projections. My intent is to recover the lost opportunity here, to hold these projections up to the light of actual data and to learn what we can from the discrepancies.

Although it is still premature to assess how well Bowen and Sosa's models will hold up over the full 25 years, 10 years out is a reasonable point to begin taking stock for two reasons. First, there is now a good deal of additional data available from the Department of Education's 1987 and 1993 National Surveys of Postsecondary Faculty (NSOPF), as well as from the National Research Council's ongoing annual Surveys of Earned Doctorates (SED) and Surveys of Degree Recipients (SDR). Bowen and Sosa made extensive use of the latter two data sources while the NSOPF results were not yet available to them. Thus, we now have access to a continuing series of data that calibrates well with Bowen and Sosa's sources as well as a rich new source that helps to extend our knowledge further.

Second, there has recently been a great deal of concern that, far from the projected shortages, the oversupply of PhDs is now as severe as ever, despite the current economic recovery (e.g., Brodie, 1995; Hartle & Galloway, 1996; Kelley, Pannapacker, & Wiltse, 1998; Schuster, 1995). Indeed, some observers have suggested that the continuing tight academic job market of the nineties has created yet another "lost generation" of scholars, like that of the seventies, who were raised on expectations of abundant faculty jobs only to be bitterly disappointed. The possibility that these new scholars may have been spurred on as graduate students by Bowen and Sosa's very projections raises the question of whether such efforts can do more harm than good. Thus, although it may seem unfair to begin judging a 25-year projection after the passage of only 10, when one considers that *Prospects for Faculty* was partly intended to fill a need for accurate data about the expected market for PhDs at the time that prospective graduate students would be likely to graduate, then the average time to complete a PhD, variously estimated at 6 to 8 years in the arts and sciences (Bowen & Rudenstine, 1992), becomes a very important milestone indeed.

This article will not be able to answer all of these questions. My plan is, first, to explain how Bowen and Sosa's model operates and what specific assumptions about higher education it employs, and second, to look at current data to explore how well the projections hold up against the performance of the academic labor market in this decade. Finally, I will try to analyze, where they have held up less well, what specific factors may have changed or what assumptions may have broken down to thwart them. Armed with this knowledge we can then begin to think about whether and how models like Bowen and Sosa's can be improved upon.

The Bowen and Sosa Model

Roughly speaking, Bowen and Sosa's faculty "universe" consisted of full-time, PhD-holding faculty teaching in arts and sciences disciplines at research, doctoral, comprehensive, and liberal arts institutions. It excluded, among others, faculty teaching at community colleges, in professional disciplines such as engineering, and all part-time and adjunct faculty. Although the group comprises only about one-sixth of all faculty in the United States today (the fraction was closer to one-fifth at the time of Bowen and Sosa's writing, excluding proprietary institutions in both cases), it still represents the "traditional core" of college and university faculty (Kirshstein, Matheson, & Jing, 1997). The main justification for this narrow scope is that Bowen and Sosa, along with much of the higher education establishment at the time, were not interested in the rest of the faculty, or rather were primarily interested in the ideal of the liberal arts professor, against which all academic careers still seem to be measured. The limitation, however, carried some distinct advantages for modeling supply and demand. The prevalence in the arts and sciences of the traditional career path, from graduate school to the tenure track to retirement, renders the labor market for these faculty, in theory, reasonably self-contained. Excluding part-timers and adjuncts (not to mention those without a PhD), reduces the complication of having to look outside this "pipeline" for sources of faculty supply. Excluding professional school faculty, who often have greater career mobility with the nonacademic world, reduces the complication of accounting for large numbers of faculty who may be more likely to leave and reenter academia. Thus, Bowen and Sosa were able to construct a labor market model with essentially just three components: two of demand and one of supply (see Fig. 1). The demand for new faculty is comprised of the need for "net new positions," or the overall growth in the size of the faculty, and the need to replace existing faculty as they "exit," either by quitting, retiring, or dying. The single component of faculty supply in the model is newly created PhDs. The predictability of the labor market by this model is ultimately dependent on the computability of these three components.

The net new position demand, for example, is assumed to be entirely determined by enrollment levels, which are projected as follows. First, the projected college-age population of the United States for the next 25 years is determined from census data. These are combined with national enrollment rates in higher education for each age group to project, assuming no change in participation rates, the overall enrollments for each year. The figures are then adjusted by the percentages of total enrollments that can be expected to be in the arts and science disciplines and in the specified types of institutions, in order to arrive at the numbers of students that the faculty will be required to teach. Finally, the expected institutional student-faculty ratios are applied to determine how many total faculty should be employed in each year. Subtracting the number of current faculty gives the number of net new positions demanded.

The replacement demand is calculated by applying known exit rates to the current population of faculty by age group. Life-expectancy tables give the percentages of faculty likely to die and retire per year at each age level. Historical data indicate the percentage expected to leave academia by choice in any given year as a function of tenure status, which is combined with the percentages of faculty with tenure in each age category to determine "quit rates" by age level. Sorting faculty into 5-year age categories for convenience, these three different percentages are applied to compute the total number of current faculty who can be expected to need replacement in any given 5-year period. Those remaining are carried into the next age category, where the cycle of calculations repeats itself for the next period. The model allows the net new positions demand to be either positive or negative, depending on whether the overall size of the faculty is growing or shrinking, while the replacement demand is assumed always to be positive and always equal to the full number of faculty exits. Thus, for example, an institution that chooses not to replace retiring faculty members is considered to have a positive replacement demand, offset by a negative net new positions demand, resulting in zero total demand, or zero actual hires.

Finally, the supply is calculated by estimating the total number of new PhDs to be awarded in any given year, then reducing this number first by the percentage of those degrees that will be awarded in the arts and sciences, second by the percentage of recipients who are U.S. residents, third by the percentage who are likely to seek jobs in academia (the "academic share"), and last by the percentage in the specified institutional sectors.

In overall structure Bowen and Sosa's model is almost identical to that pioneered by Allan Cartter (1976). It introduces significant refinements over Cartter's model, however, including a more solid grounding of the coefficients in empirical data and a relentless disaggregation of those data, such that most of the calculations are performed individually for each faculty age group, discipline, and institutional sector. Nonetheless, there are still a large number of assumptions made, and these can be viewed as falling into two types: theoretical assumptions that must be made in order to employ the model at all and numerical assumptions of the values of the coefficients within the model.

The theoretical assumptions can be further differentiated into two types: those dealing with the functioning of academic institutions as hiring and decision-making organizations, and those dealing with the functioning of academic labor markets. Among the first type, one of the main theoretical assumptions of Bowen and Sosa's model is that enrollments,

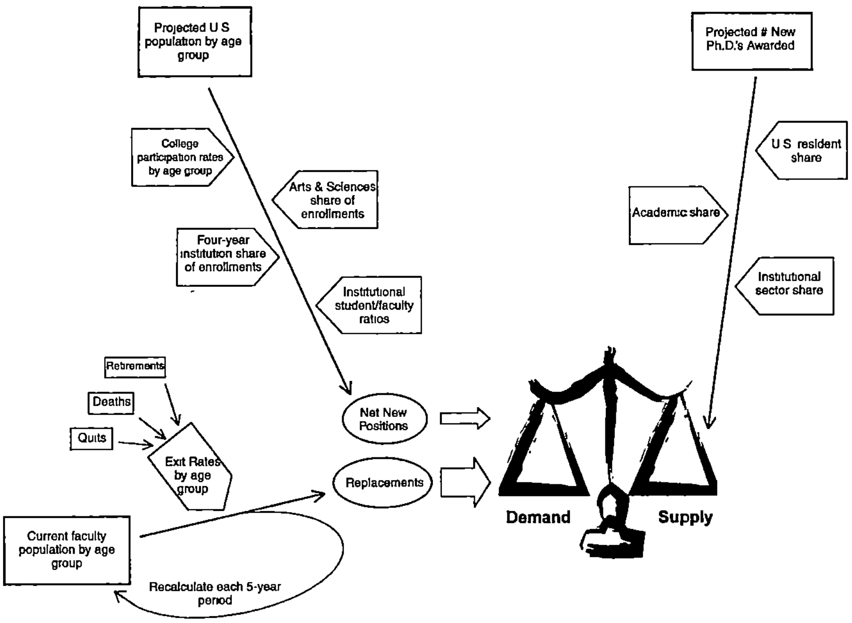

Projected U S population by age group
College participation rates by age group
Arts & Sciences share of enrollments
Four-year institution share of enrollments
Institutional student/faculty ratios
Retirements
Deaths
Quits
Exit Rates by age group
Net New Positions
Replacements
Current faculty population by age group
Recalculate each 5-year period
Projected # New Ph.D.'s Awarded
U S resident share
Academic share
Institutional sector share
Demand
Supply

FIG. 1. Bowen and Sosa's Model

tempered by student-faculty ratios, will be the sole significant determinant of faculty size in an institution. This ignores the possibility that other demands on faculty workload, such as research and service, may change, creating other pressures on faculty size. It also ignores the impacts that organizational factors can have on the ability of institutions to respond to changes in enrollments. Tenure structures and the uncapping of mandatory retirement, for example, may reduce the ability of an institution to respond to lower enrollments, but the use of part-time or nontenure-track appointments could also improve institutional responsiveness to higher enrollments. Hansen's (1985) studies, suggesting that changes in retirement legislation would prove mostly manageable for higher education institutions, provided some justification for Bowen and Sosa's decision to ignore the former factors. Ignoring the latter, however, may have been less warranted.

It is one thing to assume that faculty size responds primarily to enrollments. It is quite another thing, however, to assume that *full-time, PhD-holding* faculty size responds primarily to enrollments. Mortimer, Bagshaw, and Masland (1985) had found clear evidence that most colleges and universities were already showing signs of having developed alternative hiring practices, incorporating greater numbers of part-time, adjunct, and graduate student instructors into the faculty mix. They reported that they found full-time but nontenure-track positions in use at 65% of 4-year institutions as early as 1981 and that 25% of all faculty in those institutions were part-time in '81. These findings were certainly available to Bowen and Sosa, yet they continued to focus only on full-time faculty positions. In doing so, they may have been influenced by their own personal experiences with the elite, research university sector, where fewer alternatives to the tenure-track instructional labor force had taken hold.

Bowen and Sosa, to their credit, did acknowledge a number of organizational alternatives to the enrollments-drive-faculty assumption. One was the idea that academic institutions always "want" to have more faculty and that the number of positions will therefore be a function of the extent to which they can *afford* to hire. Budgets, especially in the public sector, are not always in synch with enrollments, and Bowen and Sosa noted that such alternate theories seemed to provide better explanations for the otherwise paradoxical phenomenon of growing arts and science faculties during the eighties when arts and science enrollments were declining. Nonetheless, Bowen and Sosa rejected the possible effects of other inputs on faculty size, primarily because enrollment level was the one quantity they felt confident in being able predict.

A key theoretical assumption of the second type, dealing with the functioning of academic labor markets, is that only newly awarded PhDs will make up the supply of new faculty. This ignores the possibility that a significant quantity of "backlog" PhDs, who were not able to get the faculty positions they desired during the years of very tight labor markets, might return to absorb some of the future demand. Cartter (1976) used an "inventory-attrition model" to incorporate precisely this backlog effect into his projections, estimating that excess PhDs in any given year would have a continuing impact on the market in the immediately following years, discounted by a factor of 20% per year, so that after five years of unemployment the market supply pressure of the academic hopeful is assumed to have vanished. Bowen and Sosa's decision to reject this option in constructing their model may have led them to understate the supply of PhDs. It may also have led them to miscalculate exits in the long run. Because each of the three exit rates is calibrated to age, feeding older "new entrants" from this backlog into the system would increasingly affect the determinations of quits, retirements, and deaths down the line.

There are many numerical assumptions built into each component of the model. In the first component, for example, the net new positions demand, Bowen and Sosa assumed that the percentage of high-school graduates who go on to college, indeed the percentage of each population age group who attend college, would remain constant. They also assumed that the percentage of college students who pursue arts and science programs, which according to their analysis had declined considerably during the 1970s and 80s, would follow one of three possible trajectories in the 90s: a slight recovery, a continuing decline, or a constant state. Each of these was computed as a distinct alternative throughout the analysis. There is a particular difficulty in these numbers, however, because the HEGIS/IPEDS data on which Bowen and Sosa relied do not record the number of enrollments by discipline, but only the number of degrees by discipline. Thus, Bowen and Sosa were forced to use the distribution of degrees as proxy for the distribution of enrollments. There is a significant amount of enrollment crossover among undergraduate degree earners, however. Depending on the type of institution, many may enroll in arts and sciences courses for up to half of their programs, even when their degrees are in professional areas such as business or education. Thus, Bowen and Sosa may have been led to overstate the declines in arts and science shares during the 1970s and 80s and, similarly, to miscalculate the trends in student-faculty ratios.

Finally, Bowen and Sosa assumed that the student-faculty ratios, which had decreased dramatically during the 80s, would remain constant, continue on a slight decline, or increase slightly (again, each of these possibilities was carried independently throughout the analysis). They explicitly rejected the possibility of large increases in student-faculty ratios, even in the case of increased enrollments. Because they rejected the possibility that the percentage of faculty working part-time might change, however, they calculated these ratios using only the numbers of full-time faculty. This numerical assumption seemed justified by some of the available evidence (most notably, Tuckman & Pickerill, 1988, who found no discernible trends in either direction for the use of part-time faculty), but unjustified, as noted above, by other evidence (Mortimer et al., 1985). It certainly led Bowen and Sosa to underestimate the range of possible changes in the student-faculty ratios.

In projecting the second component of the model, replacement demand, Bowen and Sosa assumed that rates of quits, retirements and deaths would remain constant within each age group and that no faculty would work past the age of seventy (all who had survived to that point were assumed to retire). Because quits are highly correlated to tenure status, however, this assumption carries with it the further assumption that the distribution of tenure within each age group would remain constant. This may have led Bowen and Sosa to further miscalculate quits if, as suggested above, new hires were increasingly made of older PhD holders: larger percentages of faculty being hired at older ages would effectively decrease the percentage of faculty with tenure in each age group, leading to an increase in quit rates for those groups as the model progressed through its 25-year span. By assuming that quit rates would be constant for each age group throughout the 25-year period, the Bowen and Sosa model had no way to capture this phenomenon.

In projecting the supply component of the model, Bowen and Sosa assumed that the number of new PhDs awarded in arts and science disciplines each year, which had declined by large percentages in the past decades, would level off at the 1987 levels. They also assumed that academic shares, the percentage of PhDs who seek academic careers, would continue to decline gradually. This factor, academic shares, is perhaps the most difficult to measure, because a graduate's status as a "seeker" of an academic position, as opposed to one in government or industry, is a subjective, not to mention fleeting, matter, often as much a reflection of how many jobs are being offered as of what the graduate hopes to get. What Bowen and Sosa used was the percentage of current PhD holders who have full-time faculty positions from SDR. That is, in order to project the number of new candidates for academic jobs, they used the existing academic share of old PhDs as a proxy for the intended academic share of new PhDs. Because the existing academic share, according to their analysis, had been steadily declining over the previous decade, this may have led them to overstate the intentions of the younger cohorts to enter academia.

From all of these assumptions, Bowen and Sosa arrived at four basic models they believed represented the most likely labor market scenarios for the next 25 years. Figure 2 illustrates these projections, showing the anticipated supply of faculty for each 5-year interval from 1987 to 2012, along with the total demand for each of the four models. In the first model, for example, institutions are expected to hire a total of 18,670 new faculty over the entire 5-year period between 1987 and 1992 (this is the sum of net new positions and replacements), while a total of 32,538 new academic share PhDs will be produced, resulting in an excess supply of 13,878 PhDs. It is worth noting that throughout the four models, the replacement component of demand far outweighs the net new positions component: averaged over the 25-year projections, net new positions account for only 3% to 14% of the total demand.

Despite the range of assumptions regarding enrollments and student-faculty ratios embedded in these models, the first two nonetheless projected remarkably similar academic labor markets: considerable excess supply of PhDs through no later than 1997, followed by a sharp increase in demand throughout the following decade. These two models assumed, respectively, declining arts and sciences shares of enrollments combined with declining student-faculty ratios, and constant arts and sciences enrollment shares combined with constant student-faculty ratios. The third model, which assumed constant enrollment shares and declining ratios, projected the same general pattern, but with higher demand throughout. Finally, the fourth model, involving the most sanguine of Bowen and Sosa's assumptions (rising enrollment shares only partially offset by increasing ratios), projected demand for PhDs beginning to outstrip supply no later than 1992. Taken together, the models make it easy to see the intended impact of *Prospects for Faculty*: no matter what initial assumptions we use, academic labor markets in the arts and sciences will remain tight for a relatively short period, then begin producing demand for new PhDs far in excess of supply for the foreseeable future.

Current Data

Current data allow us to test how well some of these assumptions are holding up. In all of the analyses that follow I have attempted, as much as possible, to reproduce the exact subset of the academic profession used by Bowen and Sosa.[1] This is not intended, as Bowen and Sosa also noted, to be a reflection of the relative importance of this small subset of

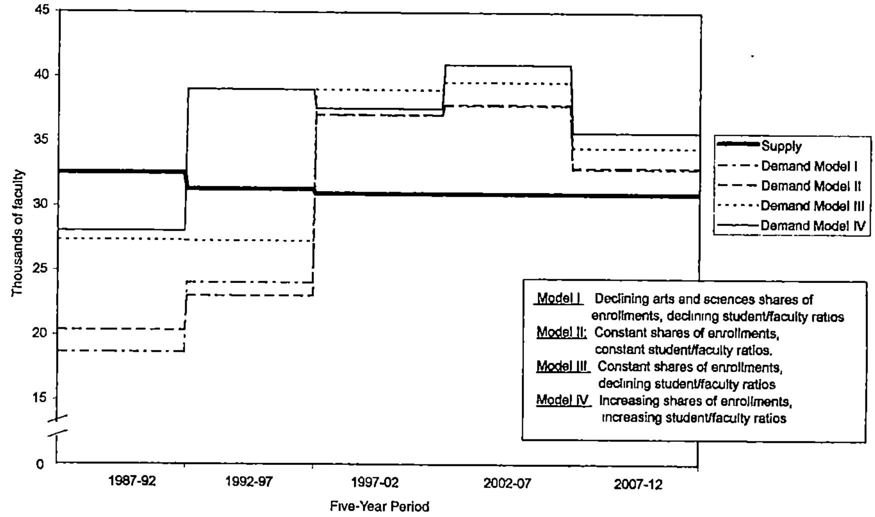


FIG. 2. Projected Total Supply and Demand for Full-time Arts and Sciences PhD Faculty

SOURCE: Bowen & Sosa (1989), Figs. 7.1–7.4.

postsecondary faculty, but rather to allow a focused look at an easily modeled segment of the market. More importantly for my purposes, it would be impossible to test Bowen and Sosa's projections without reconstructing their playing field. The danger in this approach is that today, even more so than when Bowen and Sosa were writing, the task looks increasingly like an exercise in fiddling while Rome burns. Much of the growth in faculty numbers over the past decade has been outside of the arts and sciences and, more importantly, in part-time or non-tenure-track positions and in 2-year and proprietary institutions. The continuing appeal of the traditional faculty career, however, still appears to be driving doctoral enrollments in the arts and sciences, and still appears to hold the imagination of the higher education community in its sway. For this reason it is likely that studies of Bowen and Sosa's limited faculty universe will continue, even after they have been reduced to nothing more than nostalgic lamentations of its general demise.

We know from the data, first of all, that enrollments in the arts and sciences have exceeded even the most optimistic ranges projected by Bowen and Sosa, right from the start of the decade. Figure 3 shows this trend, comparing each of the three projection scenarios (increasing shares, constant shares, and declining shares) to actual enrollments through 1995. From the 1987 baseline of 1.78 million FTEs, actual enrollments grew 22% to 2.15 million in 1995, although the highest of the three projections for that year was just 1.97 million (interpolated from the adjacent 5-year projections), an increase of less than 11%. The disparity was even greater in 1992, but appears to be diminishing if the trend of current data is extended out to 1997.

The source of this disparity is also clear. Although Bowen and Sosa projected that participation rates would remain constant and, thus, that overall enrollments in higher education would decline along with the declining national population of 18-24-year olds, the data show that participation rates increased significantly, from 57% of high-school graduates enrolling in college within twelve months of graduation in 1987, to 65% in 1996 (National Center for Education Statistics, 1998). This is shown in figure 4, which compares the projected total enrollments to the actual total enrollments. Although a major part of the growth in overall enrollments took place outside of the institutional categories of Bowen and Sosa's universe (note that the 4-year institution enrollment line in Figure 4 is not as steep as the line for all institutions), it is clear that the very first numerical assumption of *Prospects for Faculty*, the level of overall enrollments, was well off the mark in the first decade. The highest of Bowen and Sosa's three projections for arts and sciences *shares* of enrollments (the percentage of enrollments in arts and science disciplines), on the other hand, was right on the money, as illustrated in Figure 5.

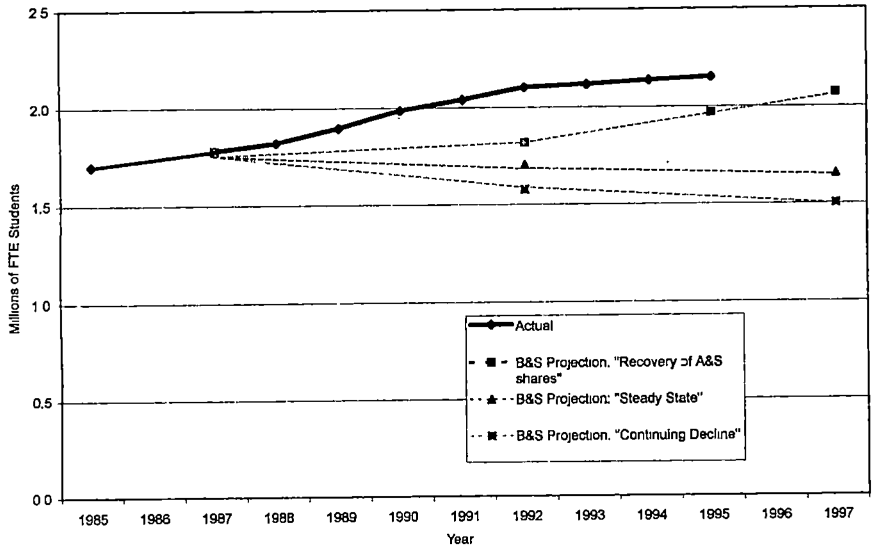


FIG. 3. Arts & Sciences FTE Enrollments in Four-Year Institutions (projections adjusted to match actual enrollments at 1987 baseline)

SOURCES: IPEDS and Bowen & Sosa (1989), Fig 4.11.

This figure shows the actual share of BAs awarded in the arts and sciences closely following Bowen and Sosa's "optimistic," or increasing shares, projection used in Model IV. Because of this, I will focus only on Model IV in the following analyses, keeping in mind that this model predicted the earliest swings in the labor market: considerable excess demand by 1992.

Despite the increase in enrollments, the number of faculty did not grow significantly over the past decade. The disparity is shown in figure 6, which compares the actual number of full-time arts and science faculty in the Bowen and Sosa universe with the numbers projected by Model IV. Faculty numbers increased by about 3% between 1987 and 1992, from 145,120 to 149,390, larger than the increase of just 0.8% projected by Model IV, but still considerably less than what the projected increase would be if we were to adjust that projection to reflect the actual enrollment growth observed above. That is, knowing that Bowen and Sosa based their projected net new positions on enrollments, if we "correct" their guess on enrollments and keep the rest of their formula intact, Model IV would have projected an increase in faculty of 12.4%, bringing the total number of full-time faculty up to 163,120 in 1992. This shows that the disparity between the projections and reality resulted from more than just the missed assumption on enrollments. In fact, it suggests that guessing correctly on enrollments would not have helped much at all to improve the accuracy of Bowen and Sosa's projections.

The difference is clearly reflected in student-faculty ratios. Bowen and Sosa were reluctant to project any increase in these ratios at all, thinking it most likely, from organizational considerations, that they would remain constant over the next 25 years, even after having dropped by over 26% in the previous decade. However, under the increasing enrollments assumption, they grudgingly allowed that it was possible for the ratios also to increase modestly, by no more than 5%. This is the figure they used in calculating the net new positions demand for Model IV, although they also allowed that the ratios could increase still further "if there were a severe financial crisis in higher education." In reality, the ratio of full-time equivalent arts and science students to full time PhD-holding arts and science faculty in 4-year institutions (the definition used by Bowen and Sosa) increased by 14.7% between 1987 and 1992, according to NSOPF and IPEDS data. In other words, although enrollments grew considerably, student-faculty ratios grew right along with them, leaving overall faculty numbers relatively flat. Thus, Bowen and Sosa's first theoretical assumption about institutional behavior, that enrollments alone would drive faculty size, completely failed to hold during the first decade of the model.

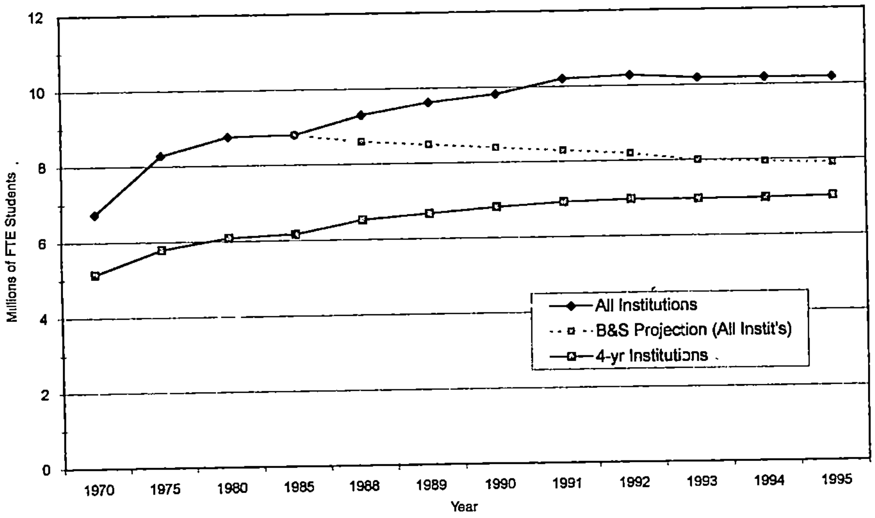


FIG. 4. Total FTE Enrollments

SOURCES: HEGIS/IPEDS and Bowen & Sosa (1989), Fig. 3.6.

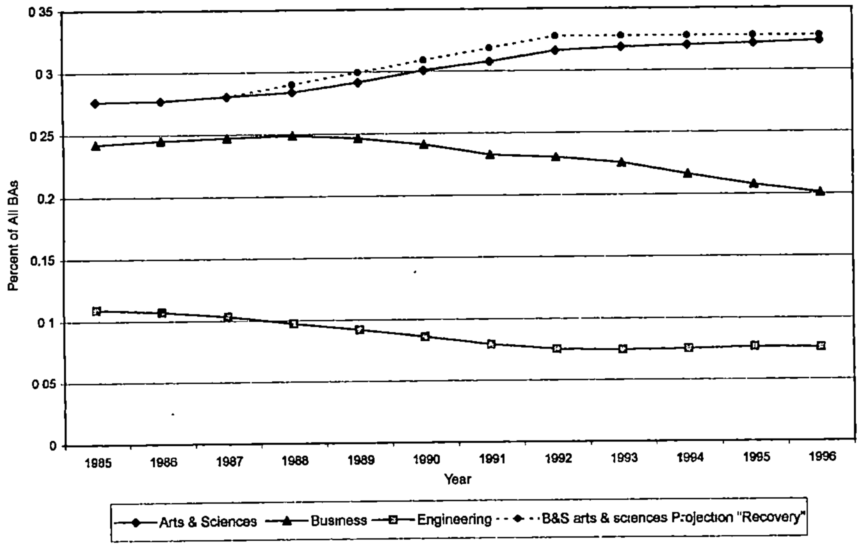


FIG. 5. Trends in Degrees Conferred: Percent of BAs at All Four-Year Institutions (projections adjusted to match actual percentages at 1985 baseline)

SOURCES: IPEDS Completion Survey and Bowen & Sosa (1989), Table 4.5.

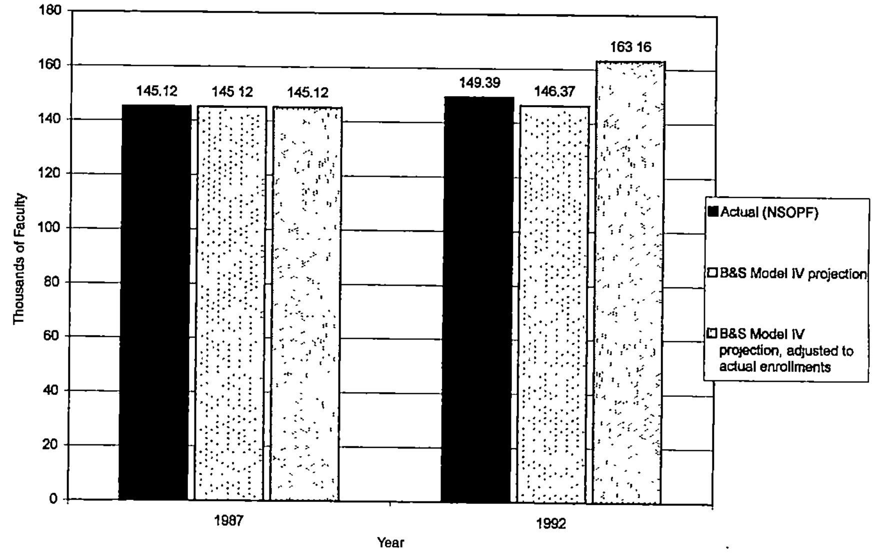


FIG. 6. Number of Full-Time Arts and Science PhD Faculty at Four-Year Institutions

SOURCES: NSOPF data, Bowen & Sosa (1989), Model IV, and author's calculations.

We turn next to the supply side of the model. Recall that all four of Bowen and Sosa's basic scenarios assumed a roughly constant production of PhDs throughout the 25-year period. The actual figures for the first decade, shown in Figure 7, are quite different. The number of doctorates awarded in arts and science disciplines grew steadily from 1987 to 1996, increasing by over 31%, from 17,460 to 22,920. The total number of PhDs awarded are only part of the picture, however. We must also estimate the percentage of these graduates who were seeking academic employment, or the academic share. For this I used the SED databases, rather than SDR, because SED has the advantage of focusing on new PhD recipients. Bowen and Sosa used the current academic share of existing PhDs as proxy for the intended academic share of new PhDs. The best proxy variable in SED, by contrast, is the percentage of new arts and science PhDs with definite employment plans in the United States who had job commitments in higher education (excluding post-docs). That is, of those doctoral recipients who had job offers at graduation, what percentage of those offers were in academia? This figure takes account of the percentage intending to stay in the United States, thus automatically correcting for the growth in the proportion of PhDs awarded to nonresidents in the past decade. Moreover, as a proxy for the academic share of job seekers this number should tend, if anything, to understate

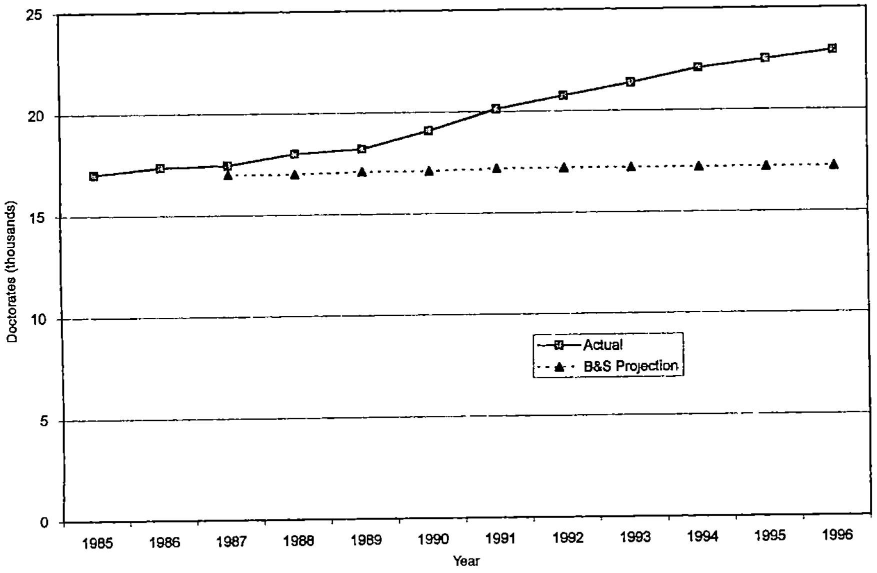


FIG. 7. Doctorates Awarded in Arts & Sciences

SOURCE: SED Doctorate Records File.

the academic share in the tight academic labor market of the past decade, as we would expect higher percentages of academic job seekers to be without offers at graduation than of nonacademic job seekers. (Recall that Bowen and Sosa's SDR proxy was more likely to have overstated the academic share of new PhDs.) The data for this quantity are shown in Figure 8, which indicates that the academic share remained roughly constant at around 50% from the mid-eighties through at least 1993, the last year for which data were available.

Applying this information to the numbers of PhDs from Figure 7 yields the top curve of Figure 9, which is the estimate of the actual number of new PhD recipients in the arts and sciences who sought faculty positions, or the observed supply of faculty. Notice the sharp increase in the slope of this curve from 1987 onward, which roughly parallels the increase in the total number of PhDs awarded (Fig. 7) reduced by the constant 50% academic share (Fig. 8). This is juxtaposed in Figure 9 with the projected faculty supply (eight thousand per year, following Bowen and Sosa's assumptions of constant PhD production and academic shares). By contrast, the lower curve of Figure 9 represents the actual number of these new PhDs hired each year (the demand), also from SED data. The gap between the number of candidates and the number of hires gives an indication of the real tightness of the academic labor mar-

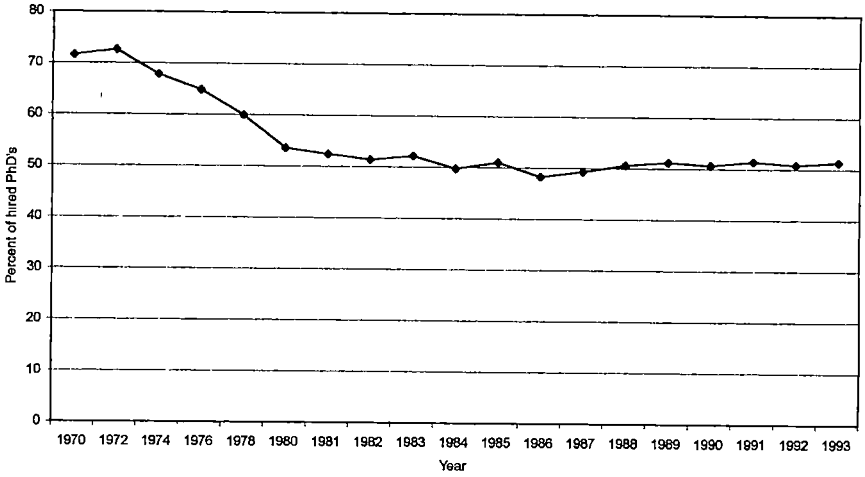


FIG. 8. Academic Share: Percent of New Arts & Science PhDs with Definite Employment Plans in the United States Whose Job Commitments Were in Higher Education

SOURCES: NCES 1995 Condition of Education, SED.

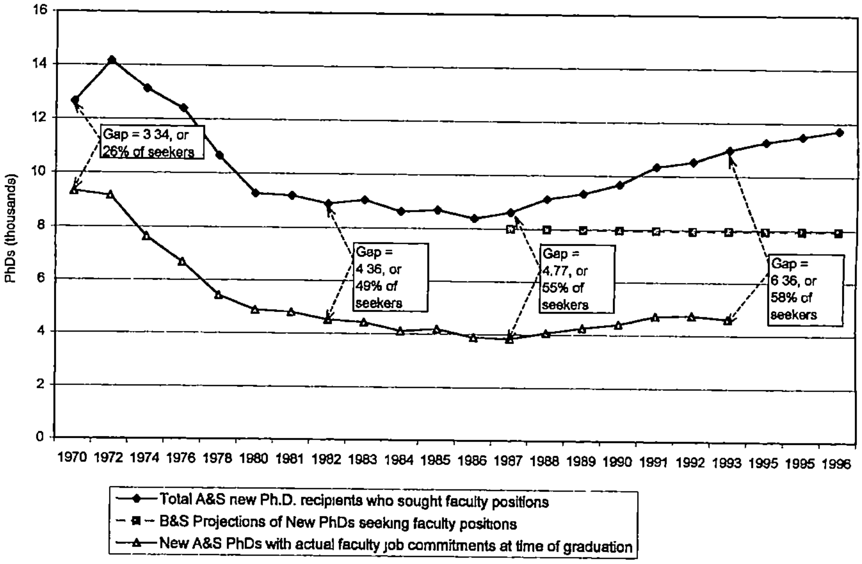


FIG. 9. Supply of New Academic PhDs in Arts & Sciences

SOURCES: SED, Bowen & Sosa (1989), Table 6.6, and author's calculations.

ket for new PhDs, increasing steadily throughout the eighties and early nineties. As indicated in the figure, these calculations suggest that 4,360 doctoral recipients, or 49% of the academic job seekers, did not find positions at the time of graduation in 1982. This gap grew to 4,770, or 55% of the candidates in 1987, and an astonishing 6,360 PhDs, or 58% of the candidates, in 1993, the last year for which data were available.

Thus, Bowen and Sosa's projections slightly underestimated the net new positions component of demand and greatly underestimated supply. What about the replacement component of demand? This quantity cannot be determined directly, because we have no data on how many faculty actually exited the profession in any given year. We turn instead to the overall balance of supply and demand, where the number of replacements can be imputed from other quantities. Table 1 compares the supply and demand data from the 5-year period 1987–1992 with the projections arising from Model IV over the same period. (Note that, up to this point, I have been using figures for all arts and science PhDs in the calculations of supply. These must now be reduced, following Bowen and Sosa's method, by the percentage expected to be hired by 4-year institutions. Thus, for example, the roughly eight thousand academic job seekers per year projected in Figure 9 becomes a total of not 40,000, but only 33,498 projected to be available for hire over the 5-year period in Table 1.) The actual net new positions number, 4,263 faculty, comes from NSOPF data (see Fig. 6), while the actual number of new PhDs hired, 21,278 faculty, comes from the sum of all new PhD job commitments in academia over the 5 years (see the lower curve of Fig. 9), multiplied by the 84% (using Bowen and Sosa's figure again) expected to go to 4-yr institutions. Finally, the "actual" number of replacements demanded, 17,015 faculty, is computed by the difference between the number of new PhDs hired and the net new positions.

The first quantity to notice in Table 1 is the huge discrepancy in excess supply, 12,866 PhDs, indicated by the difference between the actual data and the projections. This excess is not the total number of actual jobless faculty candidates during the period (18,300), but merely the total number above and beyond those who were expected to be jobless, given the projections of Model IV. This is one measure of the nineties' "crisis" in the academic labor market: the disparity between the actual number of jobless PhDs and the expected number, if one were to have believed these widely publicized projections. Interestingly, it is almost evenly split between the higher than projected number of PhDs awarded (roughly 6,100) and the lower than projected number of PhDs hired (roughly 6,800).

The second important discrepancy in Table 1 is the almost ten thousand fewer replacements actually demanded than projected. It is hard to imagine that this many fewer faculty (only 37% of the projected quantity) either quit, retired or died during this 5-year period than Bowen and Sosa's life expectancy tables and age-specific quit and retirement rates had anticipated. There simply is not that much room for error during the first 5-year period, no matter how poorly specified those coefficients are. Instead, this analysis forces us to assume that far more exits actually did take place, and that they were replaced not with new PhDs as specified in the model, but with older PhDs from the backlog of jobless candidates built up over the previous decade of excess supply. The sharp growth in postdoctoral research positions during this period (Barkume, 1996–97; Brodie, 1995) is certainly a manifestation of this phenomenon, because post-docs provide a mechanism for returning jobless PhDs into the supply pipeline after one to 3 years.

We can go further in analyzing the sources of the gap between expected net new positions and actual net new positions. First, some of the difference is made up by relaxing Bowen and Sosa's assumption of constant ratios of part-time to full-time faculty, which we know not to have held up in the nineties. Figure 10 shows that the percentage of faculty who worked part time grew from 11% in 1987 to 14% in 1992. If we convert the earlier calculations of student-faculty ratios to also take account of part-time faculty, the 14.7% increase in the ratio of FTE student enrollments to FT faculty is reduced to only a 12.6% increase when using the number of FTE faculty.[2]

TABLE 1

Supply and Demand for Arts and Science PhD Faculty in All Four-Year Institutions (Totals for Five-Year Period from 1987 to 1992)

	Actual	Model IV Projections	Difference
Net new positions	4,263	1,200	3,063
Replacements	17,015*	26,863	(9,848)
New PhDs hired (Total demand)	21,278	28,063**	(6,785)
New PhDs earned (Supply)	39,578	33,498	6,081
Excess supply	18,300	5,435	12,866
Excess as % of supply	46%	16%	

* Imputed by subtracting net new positions from the number of new PhDs hired.

** Computed by adding projected net new positions and projected replacements (exits).

Still more insight can be gained from examining changes in the age profile of faculty from 1987 to 1992. In Figure 11 we see that the overall age distribution for full-time faculty has shifted to the right, reflecting the aging process of a substantially fixed body of continuing faculty during this period. We also see, however, that the distribution has flattened somewhat, with the large spike at the 45–49 age group in 1987 (representing the expansion hires of the sixties) shrinking relative to the neighboring groups as it moves into the 50–54 age group in 1992. This illustrates the general principle that Bowen and Sosa noted in rejecting the "mass retirements" hypothesis: that non-normal distributions will become more normal over time, as the processes of attrition (quit and death rates) and replacement whittle down the spikes.

Looking more closely, we can also see that significant numbers of new hires are taking place in the 40–44 and 45–49 age groups. This is evident because these groups in 1992 are as large as, or larger than, the preceding age group was in 1987. Thus, for example, even though we know that a certain number of the faculty who were 40–44 years old in 1987 will have quit or died by 1992 (between 8% and 16%, according to Bowen and Sosa's exit assumptions), the 45–49 age group in 1992 is no smaller at all. Thus, an equivalent number of replacement hires must

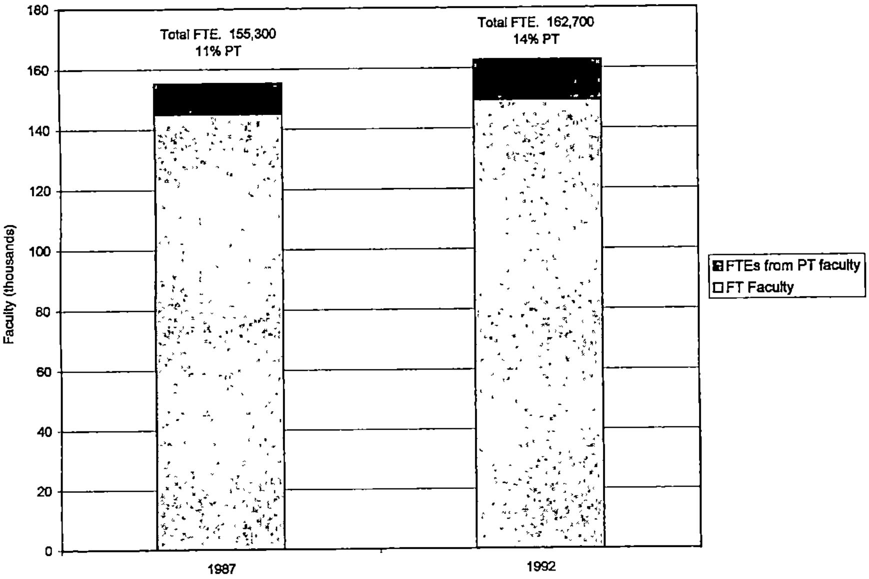


Fig. 10. FTE PhD Faculty in Arts & Sciences at Four-Year Institutions

Source: NSOPF '88, '93.

have been made directly into the 45–49 age group. These hires are not likely to have been newly awarded PhDs, however, nor even new PhDs coming off of one- or 2-year post-docs: the median age at receipt of the PhD increased only slightly during this period, from 33.8 years in 1988 to 34.2 in 1993, not enough to supply a significant increase in numbers of new PhDs in the 45–49 age group (Sanderson, Dugoni, Hoffer, & Selfa, 1999). Rather, as we deduced earlier, these must be older PhD holders who were part of the "backlog" of excess supply from the previous decade.

Finally, it is also apparent from the figure that faculty are increasingly working past age 70, and even working full-time at that age, contrary to the retirement assumptions of Bowen and Sosa. Although these numbers are not large, they are clearly significant, nearly quadrupling from just 460 faculty in 1987 to 1,700 in 1992. This phenomenon appears to defy Holden and Hansen's (1989) work, which maintained that, despite the uncapping of mandatory retirement as a result of the 1986 Age Discrimination in Employment Act (ADEA) Amendments, "there is little solid evidence that faculty members wish to work beyond age 70" (p.83).

Looking now at the age distribution of part-time faculty (Fig. 12) we also see some interesting trends. First, there are striking increases in the number of part-time faculty in each of the 45–69 age ranges from 1987 to 1992. It is possible that some of these new part-timers are former full-time faculty who have chosen to switch to part-time, but the more likely explanation, at least in the 45–59 age groups, seems to be that significant numbers of new, older PhD holders who were not previously in academia have been hired to work part-time, again in agreement with the "backlog" theory. In the 60–69 age ranges, on the other hand, it is reasonable to assume that these new part-timers represent existing full-time faculty who would otherwise have retired, but chose instead to continue on part-time. This last proposition, combined with the increase in the number of over-70 full-time faculty, represents the only real evidence we have that Bowen and Sosa's model may have overestimated the replacement demand due to retirements.

The small possible reductions in retirement rates, however, even combined with the moderate increases in part-time faculty, still only make up for a small portion of the gap in actual faculty numbers as compared to the expected numbers of Model IV, given what we know about enrollment growth in the arts and sciences. Another possible partial explanation is that the use of arts and science degrees as proxy for enrollments has caused us to overestimate the true growth in arts and science shares of enrollments. Because business and engineering degrees, for example, both declined as percentages of all BAs between 1987 and 1996 (see

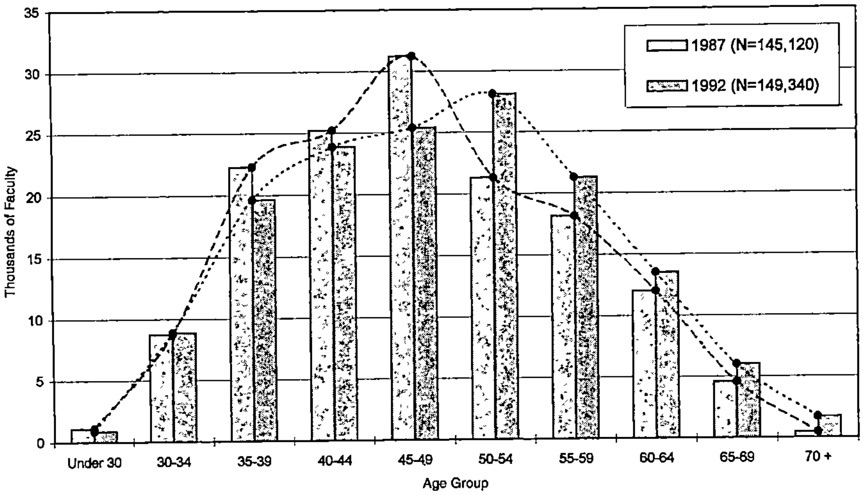


FIG. 11. Distribution of Full-Time Faculty by Age: All PhD-Holding Faculty in Arts & Sciences at Four-Year Institutions

SOURCE: NSOPF '88, '93.

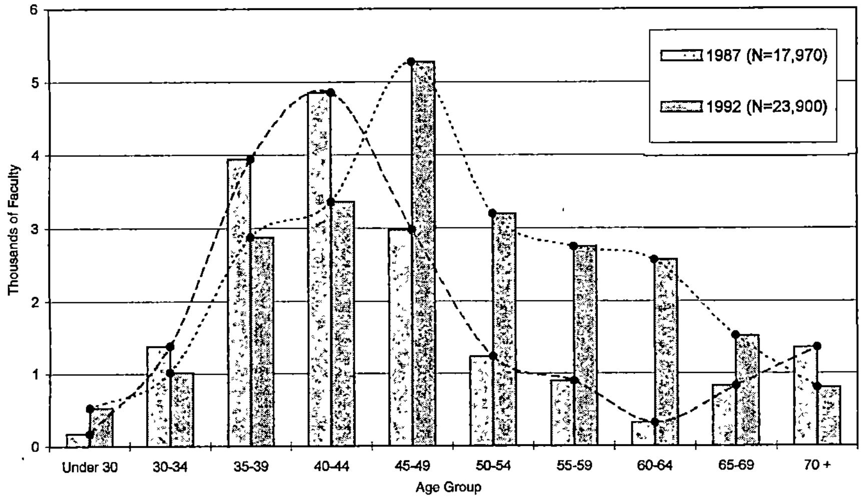


FIG. 12. Distribution of Part-Time Faculty by Age: All PhD-Holding Faculty in Arts & Sciences at Four-Year Institutions

SOURCE: NSOPF '88, '93.

Fig. 5), those portions of total arts and science *course* enrollments that were previously filled by students pursuing degrees in business and engineering would also have declined. In other words, an increase of one arts and science degree at the expense of, say, one business degree may really only be a net gain of perhaps one half of a student course load in arts and sciences if, as in some institutions, the business student takes half of her courses in the arts and sciences. This phenomenon, depending on the actual percentage of arts and sciences courses taken by non-arts and science degree seekers, may account for perhaps another percentage point off of the observed increase in student-faculty ratios. It would be significant, but still not fill the gap. The only remaining explanation to account for the 11% of increase in observed student-faculty ratios outstanding is perhaps the simplest: institutions ignored enrollment-driven demand, intentionally allowing student-faculty ratios to increase rather than hiring new faculty.

Summing up, the data have shown that Bowen and Sosa overestimated replacement demand for faculty during the first two 5-year periods, but also that they underestimated the growth in both undergraduate enrollments and PhDs awarded. In the end, the low level of replacement demand and high level of supply overwhelmed the balance of the labor market. The situation was exacerbated by the addition of a large "thumb" on the scale: the total demand for faculty was met partly by older PhDs and by part-time faculty, in defiance of the assumptions of Bowen and Sosa's model that limited their vision to new PhDs and full-time faculty only.

Implications and Discussion

As a first comment, I should state that there is clearly much more work to be done here. One of the major strengths of Bowen and Sosa's projections over those of earlier forecasters, as noted earlier, was the disaggregation of data fed into the model. Every one of its projections was broken down into detailed subsets of supply and demand for each academic discipline and each institutional type. Although I have taken pains to reproduce Bowen and Sosa's playing field in the aggregate, limitations of time and resources precluded extending my analysis to these levels. There is significantly more to be learned from doing so. Each academic discipline follows, to a certain extent, its own laws of supply and demand (see, e.g., Freeman, 1999; Goldman & Massy, 2001; Nerad & Cerny, 1999; Pollak, 1999). In fact, Burke's (1988) study found evidence of significant variation in the labor market for faculty among specialties within disciplines at research universities. Each institutional sector also tends to hire faculty differently, in accord with variations in educational mission, leading to what Youn (1988) calls "a multiplicity of overlapping markets." Caution is required here, however: although Fairweather (1995) finds some evidence for the segmentation of the faculty labor market by institutional sector, he warns that his data are also consistent with the idea of a single national market for faculty, undifferentiated by institutional type. Fairweather's results, therefore, based on quantitative analysis of national (NSOPF) data, force us to qualify several earlier depictions of stark differences among institutional sectors (e.g. Caplow & McGee, 1977; Clark, 1987; Finnegan, 1993). It may be that formerly distinct segments of the academic labor market have become more consolidated in recent years, as the excess supply of PhDs allowed institutions of diverse mission to demand higher, more uniform scholarly credentials of prospective faculty, at the same time that anxious job seekers were forced to forgo the luxury of being able to choose an institutional sector on the basis of mission. Finnegan observed both of these phenomena among the cohort of faculty hired during the oversupplied market for PhDs in the 1970s. As long as the jury is still out on this question, however, it certainly seems prudent to disaggregate data by institutional sector where feasible. Finally, there is also some evidence that disaggregation by geographical region, perhaps even more than by discipline or institutional sector, may further improve our ability to forecast academic labor markets (Chatman & Jung, 1992; McGinnis & Long, 1988).

My analysis has also raised some interesting questions that suggest a need for further data collection. For example, can we measure the actual number of quits, rather than imputing the number from other values, as I have done? If so, it would be possible to test independently the various quit rate assumptions of Bowen and Sosa's model and thus more completely assess its validity. Similarly, it would be helpful to be able to determine more accurately the sources of supply. How many new hires are in fact taken from the ranks of older PhDs, and what are the career histories of those individuals? So far the evidence is only secondhand, as in my analysis, or anecdotal (but see, e.g., Rosenfeld & Jones, 1988, for an older look at a single discipline).

Two other questions that did not come up in this analysis may also help to improve the model. The first has to do with the use of graduate student teaching assistants in research and doctoral universities. There appears to have been a significant increase in such usage in the past decade, which Goldman and Massy (2001) suggest might explain some of the reduction in hiring of full-time faculty at those institutions. Little research has attempted to document this increase, however, and the

NSOPF data offer few clues. Yet, a greater use of teaching assistants might have provided one of the escape valves that allowed student-faculty ratios to increase, as noted above, without the expected pressure on educational quality.

The second question has to do with the changing nature of students, particularly the oft-noted growth during the past two decades in the numbers of non-traditional students, older students and those enrolling part-time (Barbett & Korb, 1995; O'Brien, 1992). Edmondson (1997) suggests that new student demographics are already causing a segmentation of higher education into four distinct markets, requiring new approaches to institutional planning and staffing. It may be that the increased hiring of part-time and adjunct faculty observed in the current data is a natural institutional response to the needs of part-time and older students for more flexible course scheduling or for curricula that are less rigidly tied to traditional disciplines and degree programs. If such a link between new student demographics and new faculty hiring patterns could be established, it might be useful for refining projections of demand for PhDs, based on projections of enrollments by category.

Yet, it may be that little of this demographic shift is taking place among students in 4-year institutions at all. Recent IPEDS data show that the percentage of FTE enrollments at 4-year institutions that were made up by part-time students grew by only one percentage point, from 12% to 13%, between 1975 and 1994. The percentage peaked in 1994, and has remained at 12% from 1995 to 1997 (U.S. Department of Education, 2000). Most of the growth in part-time enrollments has occurred at 2-year institutions, outside our purview. Moreover, because all of Bowen and Sosa's estimates, as well as those presented here, employ full-time equivalent (FTE) enrollment figures, the part-time trend has no direct effect on our interpretations of the model and data. Did Bowen and Sosa's underestimation of the college participation rates of older segments of the population contribute to their overall underprojection of liberal arts enrollments? Certainly, but no more than did their underestimation of the participation rates of younger segments. As we have seen, the enrollment projections were the source only of numerical, not theoretical flaws in the model, and relatively blameless ones at that.

Another comment that should be made at the outset is that it is easy to find fault in hindsight, much more difficult to improve on the actual performance the next time around. I have been able to point out the glaring failures in the theoretical and numerical assumptions made in Bowen and Sosa's model, but this stops far short of offering alternative assumptions that could have been employed more effectively at the time. This is because, for the most part, theoretical assumptions are made for the sake of computability, not out of simplemindedness, while numerical assumptions are made by rarely more sophisticated methods than extending or reversing current trends, and thus are inevitably more a matter of luck than skill. The danger of all mathematical models of complex systems like labor markets is that their construction forces us to favor the predictive power of the computable factors (such as enrollments) over the unknowable ones (such as state budgets), even though past experience and theory both argue for the reverse.

Bowen and Sosa did not believe, for example, that older PhD holders never returned to academia, or that faculty size would be completely determined by enrollments alone. Rather, they calculated that other contributing factors were likely to be either too small to make a significant difference, or else too complex to be modeled feasibly. Moreover, their numerical assumptions, such as "the production of new Ph.D.'s will continue unchanged at the 1987 levels," were not attempts at prophesy, but simply admissions that they had no reasonable basis on which to make any other assumptions about the future value of the quantity. The only fair question, therefore, is not "what did they do wrong" but rather, what could we do differently if we were to attempt a new projection today? Undoubtedly, we would fiddle around the edges a bit, tweaking this or that computation or coefficient, but still be prey, in the end, to the same uncertainties.

Indeed, the foregoing analysis does indicate some ways to increase the sensitivity of the model. One of these, for example, would be to improve the validity of the use of degrees by discipline as proxy for enrollments by discipline, thereby increasing the accuracy of the calculations of student-faculty ratios. This could be done by measuring the actual percentages of course enrollments by discipline for each of the different categories of degree seekers at different types of institutions. Such a task would involve a considerable amount of transcript analysis, but would yield a simple matrix of coefficients for converting the numbers of degrees by discipline into estimates of course enrollments by discipline. An easier, and perhaps no less effective method, would be to follow the example of Goldman and Massy's (2001) model of supply and demand for science and engineering doctorates. Goldman and Massy used a measure of the total number of undergraduate degrees awarded by the institutions, distinct from the number of degrees awarded in each discipline of interest, as an indicator of the demand for teaching created by the general-education requirements of students whose degrees would eventually be in other areas. This measure, combined with the direct measure of degrees awarded in the disciplines, would help to give a fuller picture of actual teaching demands generated by overall enrollments both inside and outside the liberal arts.

Another, similar improvement would be to eliminate the assumption that the distribution of tenure would remain constant within each age group for the purpose of calculating quit rates. One could develop a new set of quit rates that took into account both age and tenure status, or age and the number of years at the institution, to allow the quit projections to respond to changes in the average age of new hires. Another adjustment would be to seek a more accurate measure of the intent to pursue an academic career, to replace the "percentage of new PhDs with definite employment plans in the United States who had job commitments in higher education" as proxy for the academic share of new PhDs. This improvement could involve a refinement of items on the SED designed to indicate labor force expectations; combining it with a model of the recent job placement records for PhDs within individual disciplines, disaggregated by institutional sectors, might further refine such a measure. This may prove the most difficult, however. For, as Ehrenberg (1991) amply demonstrates, there is a great deal that we still do not know about the career choices and expectations of doctoral students.

Although each of the adjustments suggested above would improve the model by enhancing the accuracy of its numerical assumptions, the more important conclusions to be drawn from this analysis are those that point to the relative importance of the main components of the model and its vulnerability to the failure of its larger, theoretical assumptions. The labor market data show, for example, that replacement demand is the single most important factor to get right in projecting the balance of supply and demand for faculty. This component of demand, as Bowen and Sosa themselves emphasized and the actual data confirmed, overwhelms the net new positions demand. More importantly, it also overwhelms fluctuations in enrollments. The analysis has shown that even guessing correctly on the dramatic growth of arts and science enrollments would not have contributed much to the accuracy of Bowen and Sosa's projections. The reason is that, contrary to Bowen and Sosa's theoretical assumption about the functioning of academic institutions as hiring and decision-making organizations, colleges and universities clearly acted on factors other than enrollments when making decisions about replacement and hiring of full-time faculty.

The labor market data also show that Bowen and Sosa's theoretical assumptions about the functioning of academic labor markets introduce further weaknesses into the model. Contrary to the assumption that only newly awarded PhDs will make up the supply of new faculty, we saw that significant numbers of new hires were made from the ranks of older PhD holders. Thus, it is clear that some attempt to estimate and account for the backlog of older PhD recipients who hold nonacademic positions now, but may seek academic positions if openings arose, would significantly improve the model's ability to predict faculty supply. Future models of academic labor markets would certainly benefit from the use of an inventory-attrition model like that employed by Cartter (1976, and described above). More immediately, the analysis presented here suggests that policymakers concerned with the optimal size of graduate programs should consider not just the number of expected openings in a discipline 6 or 7 years hence, but also the backlog of current PhD holders who are likely to compete for those openings.

A final example of a possible improvement to the model comes from Goldman and Massy's (2001) study. In contrast to Bowen and Sosa's fixed-coefficient model of the entire national system for arts and sciences, Goldman and Massy constructed a dynamic simulation of labor market and institutional behavior focusing on individual academic departments in science and engineering. As such, their work is not, technically, a projection of future supply and demand so much as an attempt to understand the dynamics of the major forces influencing the labor market in the past (during the 1980s). Using Markov simulation methods, it models some of the important feedback loops and market adjustments that Bowen and Sosa had to ignore. For example, classical labor economics suggests that a depressed job market for PhDs in a specific discipline should result in fewer new applicants to those PhD programs, with the effect of reducing the imbalance after a certain time-lag (Freeman, 1989). Goldman and Massy modeled this effect but found it to be too small to significantly reduce the imbalances they identified. They also modeled the function of departmental choice in optimizing the mix among faculty, post-docs and doctoral students to meet the research and teaching needs of the department, subject to the constraints of research budgets and the number of applicants to doctoral programs. This enabled them to account for the dual role of doctoral students both as additions to the future supply of faculty (as PhDs) and as reductions to the present demand for faculty (as teaching and research assistants), a significant advantage of the dynamical systems approach over the fixed-coefficient model. Goldman and Massy's working hypothesis was that doctoral program size, and hence the eventual supply of PhDs, is determined not by any output labor market considerations, but rather by departmental needs for RAs and TAs, limited only by the financial constraints of supporting the doctoral students. Their simulation results, which confirmed this hypothesis, could be used to improve further the theoretical assumptions of Bowen and Sosa's model. They clearly urge the inclusion of research funding, as well as enrollments, among the drivers of faculty hiring.

Unfortunately, as a practical matter, there may be no accurate way to project such budgets. Moreover, it is not clear how well Goldman and Massy's system could be adapted to making projections across the liberal arts disciplines. In most of the Humanities, for example, the linkages they identified among research budgets, faculty hiring, and doctoral student admissions probably do not apply. Goldman and Massy's recognition, however, that doctoral students supply a significant part of the teaching needs of a department, reducing the demand for PhD faculty, could be used to improve the accuracy Bowen and Sosa's model. Indeed, the analysis presented here suggests that expanding Bowen and Sosa's concept of "faculty" to include part-time and nontenure-track appointments, as well as TAs, would be the most significant single improvement to make in the model. Unfortunately, there is currently no good measure of either the number of TAs employed or the amount of teaching they perform in U.S. universities, with which to implement such an improvement.

This sort of discussion begs the question, however, of whether such massive forecasting efforts as Bowen and Sosa conducted are ultimately worth doing at all, when they so often seem to come out wrong. Bowen and Sosa themselves maintained that their projections would never be observed, because individuals, institutions and markets would all make adjustments in the intervening years. Yet, it is surely too facile to cite Cartter's (1976) self-exonerating claim that a good projection *should* turn out to be a bad prediction by virtue of its success at inducing responsible actors to address the problem at hand, because a wrong projection can be worse than no projection at all if it is so severe or alarmist that the response turns out to be worse than the cause. This would certainly seem to be the case if we were to credit Bowen and Sosa's projection with having lured tens of thousands of the nation's brightest students into lengthy doctoral programs on the promise of plentiful jobs in academia, only to leave them with dashed expectations, embittered outlooks, part-time employment, and many thousands of dollars of debt. Rubenstein (1999) goes so far as to accuse Bowen and Sosa, among others, of acting in bad faith by waging a "skills-shortage campaign," whose hidden agenda was access to cheap labor for industry and academe but whose result has been a wasteful overproduction of PhDs and underconsumption of scholarship. I would not venture so far as Rubenstein, but I am not sure that we would not have been better off today without *Prospects for Faculty*. The question is not without immediate relevance, for if we follow the pattern of major labor market projections in recent decades, we are already due for the next one. Given the lessons learned, I wonder whether we should blithely forge ahead with another, even given the refinements I have recommended.

I suggest that we must. There is a definite need for projections that would enable actors on both sides of the market, institutions and prospective faculty alike, to plan rationally for their futures. Insinuations of bad faith aside, the real danger seems to lie not so much in being wrong, but in being taken too literally. Bowen and Sosa recognized all too well how easy it is to lull readers into complacency with the reassuring precision of numbers, charts and tables. One cannot say often enough, it seems, that the product is merely a hypothetical model based on projections of existing trends, not a prediction of what will actually occur. It may be, therefore, that the best way to improve such models is to do a better job of laying every assumption bare to the reader. To make all of the technical refinements and improvements one can, certainly, but also to make them explicit, and adjustable on the fly. One could envision, for example, the next edition of *Prospects for Faculty* being published as an interactive web-based system, in which the user would have the ability to manipulate independently each of the coefficients in the calculations and watch the resulting effects on the projections instantly take shape. This would have the doubly beneficial effect of, first, allowing each numerical assumption to be adjusted as new data become available and second, allowing users to explore their own "what if" scenarios to see just how speculative the entire exercise truly is.

Notes

[1] Reconstructing the playing field of *Prospects for Faculty* has been a relatively straightforward task when using the same series of datasets and surveys that Bowen and Sosa used, namely the National Research Council's annual SDR and SED data and National Center for Education Statistics' (NCES) IPEDS data. Exact correspondences were sometimes more elusive when working with data from the two NSOPF studies, which nonetheless provide much richer snapshots of faculty characteristics. In those cases where the differences were small enough as to be meaningless in the context of comparing relative trends in quantities over time, minor correction factors were introduced into the data in order to match the latest Bowen and Sosa figures (usually 1987) with those from the same year in the current datasets. In all extractions of data from national surveys, weighted estimates were used where appropriate.

[2] These calculations assume that one part-time faculty member is equivalent to 0.56 full-time faculty members, a factor derived from the ratio of the number of courses taught by part-time vs. full-time faculty in NSOPF '93.

References

Barbett, S. F., & Korb, R. A. (1995). *Enrollment in higher education: Fall 1984 through Fall 1993* (NCES 1995–04–00).

Barkume, M. (1996–97, Winter). The job market for Ph.D.'s: Two views. *Occupational Outlook Quarterly, 40,* 2–15.

Blum, D. (1991, February 20). Many studies of future academic job market are said to be of little use to policy makers. *Chronicle of Higher Education, 37,* pp. A15, A19.

Bowen, H. R., & Schuster, J. H. (1986). *American professors: A national resource imperiled.* New York: Oxford University Press.

Bowen, W. G., & Rudenstine, N. L. (1992). *In pursuit of the PhD.* Princeton: Princeton University Press.

Bowen, W. G., & Sosa, J. A. (1989). *Prospects for faculty in the arts and sciences: A study of factors affecting demand and supply, 1987 to 2012.* Princeton: Princeton University Press.

Brodie, J. M. (1995, January-February). Whatever happened to the job boom? *Academe, 81,* 12–15.

Burke, D. L. (1988). *The new academic marketplace.* Westport, CT: Greenwood Press.

Caplow, T., & McGee, R. J. (1977). *The academic marketplace* (Reprint of 1958 ed.). New York: Arrow Press.

Cartter, A. M. (1976). *Ph.D.'s and the academic labor market: A report prepared for the Carnegie Commission on Higher Education.* New York: McGraw-Hill.

Chatman, S., & Jung, L. (1992). Concern about forecasts of national faculty shortages and the importance of local studies. *Research in Higher Education, 33,* 31–57.

Clark, B. R. (1987). *The academic life: Small worlds, different worlds* . Princeton: Carnegie Foundation for the Advancement of Teaching.

Edmondson, B. (1997, March). Keeping up with demographic change: Higher education at the start of the next century. *College Board Review, 180,* 24–30.

Ehrenberg, R. G. (1991). Academic labor supply. In C. T. Clotfelter, R. G. Ehrenberg, M. Getz, & J. J. Siegfried (Eds.), *Economic challenges in higher education* (pp. 143–258). Chicago: University of Chicago Press.

El-Khawas, E. (1990). *Campus trends, 1990* (Higher Education Panel Report No. 80). Washington, DC: American Council on Education.

El-Khawas, E. (1991). *Campus trends, 1991* (Higher Education Panel Report No. 81). Washington, DC: American Council on Education.

Fairweather, J. S. (1995). Myths and realities of academic labor markets. *Economics of Education Review, 14,* 179–192.

Finnegan, D. E. (1993). Segmentation in the academic labor market: Hiring cohorts in comprehensive universities. *Journal of Higher Education, 64,* 621–656.

Freeman, R. B. (1989). *Labor markets in action: Essays in empirical economics.* Cambridge: Harvard University Press.

Freeman, R. B. (1999). It's better being an economist (but don't tell anyone). *Journal of Economic Perspectives, 13*(3), 139–145.

Gill, J. I. (1992). *Bringing into focus the factors affecting faculty supply and demand: A primer for higher education and state policymakers.* Boulder, CO: Western Interstate Commission for Higher Education.

Goldman, C. A., & Massy, W. F. (2001). *The PhD Factory: Training and employment of science and engineering doctorates in the United States.* Bolton, MA: Anker Publishing.

Hansen, W. L. (1985). Changing demography of faculty in higher education. In S. M. Clark & D. R. Lewis (Eds.), *Faculty vitality and institutional productivity: Critical perspectives for higher education* (pp. 27–54). New York: Teachers College Press.

Hartle, T. W., & Galloway, F. J. (1996, September/October). Too many PhDs? Too many MDs? *Change, 28,* 26–33.

Holden, K. C., & Hansen, W. L. (Eds.). (1989). *The end of mandatory retirement: Effects on higher education.* San Francisco: Jossey-Bass.

Kelley, M. R., Pannapacker, W., & Wiltse, E. (1998, December 18). Scholarly associations must face the true causes of the academic job crisis [Opinion]. *Chronicle of Higher Education, 45,* p. B4.

Kirshstein, R. J., Matheson, N., & Jing, Z. (1997). *Instructional faculty and staff in higher education institutions: Fall 1987 and Fall 1992* (Statistical Analysis Report): National Center for Education Statistics, US Department of Education.

Lozier, G. G., & Dooris, M. J. (1987, November). *Is higher education confronting faculty shortages?* Paper presented at the annual meeting of the Association for the Study of Higher Education, Baltimore.

McGinnis, R., & Long, J. S. (1988). Entry into academia: Effects of stratification, geography and ecology. In D. W. Breneman & T. I. K. Youn (Eds.), *Academic labor markets and careers* (pp. 28–51). New York: The Falmer Press.

McGuire, M. D., & Price, J. A. (1989). *Faculty replacement needs for the next 15 years: A simulated attrition model.* Paper presented at the 29th annual forum of the Association for Institutional Research, Baltimore.

Mortimer, K. P., Bagshaw, M., & Masland, A. T. (1985). *Flexibility in academic staffing: Effective policies and practices* (ASHE-ERIC Higher Education Report No. 1). Washington, DC: Association for the Study of Higher Education.

National Center for Education Statistics. (1998). *The condition of education, 1998.*

National Center for Education Statistics DAS data access system [http://nces.ed.gov/das/]:

National Science Foundation WebCASPAR data access system [http://caspar.nsf.gov]:

Nerad, M., & Cerny, J. (1999). Postdoctoral patterns, career advancement, and problems. *Science, 285,* 1533–1535.

O'Brien, E. M. (1992). Part-time enrollment: Trends and issues. *ACE Research Briefs, 3*(8).

Pollak, M. (1999, September 15). Ph.D.'s fault universities' roles in helping them to find work. *New York Times,* p. B10.

Rosenfeld, R. A., & Jones, J. A. (1988). Exit and re-entry in higher education. In D. W. Breneman & T. I. K. Youn (Eds.), *Academic labor markets and careers* (pp. 74–97). New York: The Falmer Press.

Rubenstein, E. S. (1999, Fall). Piled Higher and deeper: The alleged shortage of highly educated workers in the U.S. is a myth. *American Outlook,* 44–47.

Sanderson, A., Dugoni, B., Hoffer, T., & Selfa, L. (1999). *Doctorate recipients from United States universities: Summary report 1998.* Chicago: National Opinion Research Center (The report gives the results of data collected in the Survey of Earned Doctorates, conducted for five Federal agencies, NSF, NIH, NEH, USED, and USDA, by NORC.).

Schuster, J. H. (1995, Fall). Whither the faculty? The changing academic labor market. *Educational Record,* 28–33.

Tuckman, H. P., & Pickerill, K. M. (1988). Part-time faculty and part-time academic careers. In D. W. Breneman & T. I. K. Youn (Eds.), *Academic labor markets and careers* (pp. 98–113). New York: The Falmer Press.

U.S. Department of Education. (2000). *Digest of Education Statistics, 1999.* Washington, DC: National Center for Education Statistics.

Youn, T. I. K. (1988). Studies of academic labor markets and careers: A historical review. In D. W. Breneman & T. I. K. Youn (Eds.), *Academic labor markets and careers.* New York: The Falmer Press.

Lamont Flowers
Steven J. Osterlind
Ernest T. Pascarella
Christopher T. Pierson

How Much Do Students Learn in College?

Cross-Sectional Estimates Using the College BASE

Introduction

What students learn during the undergraduate years in postsecondary education has been the focus of a large body of inquiry for the past 30 years (e.g., Bowen, 1977; Gardner, 1994; Kuh, Douglas, Lund, & Ramin-Gyurnek, 1994; Osterlind, 1997; Pascarella & Terenzini, 1991; Stage, Muller, Kinzie, & Simmons, 1998). The vast majority of these studies document the change, or growth, that occurs between the first and subsequent years of college. Overall, this research has taken three different forms. The first form is longitudinal (or panel) studies that trace changes in the same students over time on some standardized, objective measure of cognitive development or knowledge (e.g., critical thinking, mathematical reasoning, literacy, history, science reasoning). The average change between the freshman and senior year of college, for example, might then be an estimate of the impact of college. Recent examples of studies of this type are by Baxter Magolda (1990, 1991), Giancarlo and Facione (1997), Hart, Rickards, and Mentkowski (1995), King and Kitchner (1994), Kube and Thorndike (1991), May (1990), Saucier (1995), Thorndike, Andrieu-Parker, and Kube (1990), Thorndike and Andrieu-Parker (1992), and Zhang and RiCharde (1999).

The second form is cross-sectional (or cohort) studies that administer

Lamont Flowers is assistant professor in the Department of Educational Leadership, Policy, and Foundations at the University of Florida; Steven J. Osterlind is professor of educational psychology at the University of Missouri; Ernest T. Pascarella is the Mary Louise Petersen Chair in Higher Education at the University of Iowa; and Christopher T. Pierson is a doctoral student in higher education and law student at the University of Iowa. (Authorship is in alphabetical order.)

The Journal of Higher Education, Vol. 72, No. 5 (September/October 2001)
Copyright © 2001 by The Ohio State University

some standardized objective measure of cognitive development or knowledge to students with different levels of exposure to postsecondary education. For example, one might administer the same instrument at the same time to first-year students and seniors and assume that the difference between the average scores of the two groups is an estimate of the impact of college. Recent examples of this type of study are found in ACT (1993), Durham, Hays, and Martinez (1994), Evans (1989), Hill (1995), Jehng, Johnson, and Anderson (1993), King and Kitchener (1994), McDonough (1997), Osterlind (1997), Pearson and Rogers (1998), and Wood (1997).

The third form of study employs students' retrospective self-reports of how much they change or gain along various dimensions of knowledge or cognitive development during college. The assumption, not without some empirical support (e.g., Anaya, 1999a, 1999b; Bradburn & Sudman, 1988; Trusheim, 1994), is that students will be both truthful and accurate in responding to such instruments. Recent examples of this type of study are by Bauer (1996, 1998), Dollar (1991), Feldman (1994), Lincoln (1991), Tan (1995), and Williams (1996).

Each of these three popular forms of research provides legitimate approaches to understanding the change and growth in knowledge and cognitive development that occur during college, and in the clear majority of cases they report statistically significant and sometimes sizable indications of growth or change. Yet, if one is interested in estimating the net impact of college (i.e., how much of what happens is attributable to the experience of college and not to coincident or competing influences, or measurement error), then each of these forms of research is also limited. For example, longitudinal panel studies of change during college have the problem of not knowing whether similar students who do not attend college might also change to the same or to a similar extent. Additionally, there is the problem of the practice effect (part of the change from the first to subsequent testings may be due simply to students taking the test more than once and not to the impact of college).

Cross-sectional or cohort studies have their own methodological problems too. Probably the most important as an estimate of the impact of college is concerned with sample mortality. Because of sample mortality, seniors might represent a more academically select and motivated group than first-year students. Thus, higher average scores by seniors versus first-year students, for example, might reflect comparison group differences in academic aptitude and motivation, not simply the impact of college.

Finally, student self-reports of cognitive gains or growth during college are also limited as an estimate of the net impact of college. Probably the major issue in the use of self-reports is their veracity and the question of what it is that they measure. Unfortunately, the correlations between student self-reports and scores on objective, standardized measures are often quite modest in magnitude (e.g., Osterlind, Robinson, & Nickens, 1997; Pascarella & Terenzini, 1991; Pike 1995, 1996; Steel & Nichols, 1992). Although having low correlations with objective, standardized measures of student learning does not necessarily mean that self-reports are psychometrically invalid measures of college impact, this fact might well lead to serious questions about their content or face validity by faculty and administrators. Another important problem centers on disentangling students' self-reports of gains during college from the very characteristics of the students reporting them. For example, if seniors at a selective, residential liberal arts college report greater cognitive gains during college than seniors at a large, state university, the temptation might be to conclude that the former provides a more cognitively impactful environment than the latter. However, students enrolling in selective, residential liberal arts colleges may simply be more open to the cognitive impacts of postsecondary education to begin with than their counterparts enrolling in a large state university. Thus, the differences observed between institutional types might simply reflect differences in the characteristics of the students enrolled and not the unique impact of the institution itself (Kuh, 1993; Pascarella & Terenzini, 1991). Without taking such student background characteristics into account, it is problematic to attribute differences in self-reported cognitive gains by students with different amounts of exposure to college (e.g., sophomores vs. seniors) to the impact of college (Pascarella & Terenzini, 1991).

To be sure, there is a limited body of research that specifically attempts to estimate the net impact of college on learning or cognitive development by introducing controls for important student background characteristics (e.g., Flowers & Pascarella, 1999; Kitchener, 1992; Myerson, Rank, Raines, & Schnitzler, 1998; Pascarella, 1989; Pascarella, Bohr, Nora, & Terenzini, 1996; Rykiel, 1995; Smith-Sanders & Twale, 1997, 1998; Whitmire & Lawrence, 1996; Wolfle, 1983). The results of these studies clearly suggest that, even in the presence of statistical controls for important student background characteristics (e.g., academic ability, motivation, race, sex, high-school performance) students with more exposure to postsecondary education make greater gains in measures of learning or cognitive development than students with less or no exposure.

However, this small body of research also has a number of distinct limitations. First, with a few exceptions there is little evidence concerning the actual magnitude or size of the net impacts of college. Pascarella

(1989), and Pascarella, et al. (1996) analyzing data from two independent samples report a net advantage in critical thinking of about 0.4 of a standard deviation, but this estimate is limited to the impact of exposure to the first year of college. A second and related problem is that there is little or no information about the timing of the net impacts of college on learning or cognitive growth. Does most of the impact occur early during one's exposure to postsecondary education, or are the impacts spread more evenly across the undergraduate experience? Finally, there is little evidence on whether the net impacts of college on learning or cognitive development are general or conditional. That is, are the impacts essentially similar in magnitude for all students (i.e., general effects) or do they differ in magnitude for students with different background characteristics (e.g., academic ability, race, gender), or for students in different institutional contexts (e.g., the average academic ability of the student body)? The latter are termed conditional effects.

This study sought to address these problems in the existing research by analyzing a large, multi-institutional sample of freshmen, sophomores, juniors, and seniors who completed the *College Basic Academic Subjects Examination* (hereafter CBASE) (Osterlind, 1997). The study had three specific purposes. First, it sought to estimate the magnitude of the effects of year in school on CBASE scores with controls for important individual student precollege characteristics (e.g., academic ability, race, gender), as well as the context of the institution in which the student was enrolled (e.g., an estimate of average academic ability of the institution's student body). Second, by estimating the relative net effects of year in school, the analyses yielded evidence with regard to when the largest effects occurred in a student's undergraduate experience. Third, the study sought to determine the presence of conditional effects of year in college on CBASE scores. That is, do the net effects of college differ in magnitude for students with different precollege characteristics, or for students in different institutional contexts?

Methodology

Initial Sample

The data for the study came from undergraduate students enrolled in 56 four-year colleges and universities located in 13 states. The institutions were both public and private and consisted of liberal arts colleges, regional colleges and universities, and research universities. About 25% of the institutions were doctorate-granting universities. The students at each institution came from all classes, freshman through senior, and agreed to participate in an assessment conducted by their institution.

Thus, there was no random sampling strategy used to form the sample. All students were tested between 1993 and 1998, although there were only a small percentage in the years 1993 and 1998; hence the vast majority were tested between 1994 and 1997. Most participants in the sample were full-time students between the ages of 19 to 22, and as a group had a racial composition that approximated the national distribution for four-year institutions. Thus, despite the fact that the sample was not scientifically drawn, it nevertheless had reasonable comparability with a national profile of full-time, traditional-age college students.

Instrument

The measure of student learning used in the study was the *College Basic Academic Subjects Examination* or CBASE (Osterlind, 1997). The CBASE is a standardized measure of scholastic achievement in four content areas: English, mathematics, science, and social studies. In addition, the test assesses three reasoning competencies across the disciplines: interpretive reasoning, strategic reasoning, and adaptive reasoning. The model underlying these cognitive–processing competencies is based on the theoretical work of Hannah and Michaelis (1997). Organizationally, the CBASE is divided into four sections; each one devoted to a particular content subject. The content subjects themselves are divided into two or three "clusters" of skills for that area. The English test is composed of the two clusters reading and literature, and writing; the mathematics test is composed of the three clusters general mathematics proficiency, algebra, and geometry; the science test is made up of the two clusters laboratory and field work, and fundamental concepts; and the two clusters history and social sciences compose the social studies test. In all, nine clusters aggregate into the four CBASE subject tests. The clusters are themselves further subdivided into a number of "enabling subskills." For example, reading critically, reading analytically, and understanding literature are enabling subskills within the cluster called "reading and literature." Similarly, evaluating algebraic and numerical expressions, and solving equations and inequalities are enabling subskills within the cluster called "algebra." In total the exam includes four subject areas, nine clusters, and 23 skills.

The main focus of the test is on CBASE scores in the four content areas (i.e., English, mathematics, science, and social studies) plus a Composite score that combined all four subject tests. Evidence for valid score interpretation yielded by the CBASE is evident through a number of sources, following guidelines set forth in the Standards for Educational and Psychological Testing (American Psychological Association, 1985). Perhaps most important is information pertinent to the content validity of the instrument. The CBASE has a written rationale and a detailed plan for development. Extensive test specifications exist. These test specifications were prepared over a period of one year with advice from more than 100 content experts, primarily from universities throughout the country. Separate studies with appropriate college-level populations were conducted to determine congruence with other measures (e.g., the ACT and SAT) as well as for scaling and test standardization. The four content areas were developed through factor analysis and item analysis for individual test items (including bias and differential performance). According to independent studies of the use of the four content area scores, the CBASE can be employed to assess student knowledge and skills and to evaluate the quality and effectiveness of academic programs (Pike, 1991, 1992). In the present study the correlations between the four CBASE subject scores and students' cumulative college grades ranged from 0.22 to 0.30. The correlation between the Composite score and cumulative grades in college was 0.32. The alpha, internal consistency reliabilities for the four CBASE tests are: English 0.77, mathematics 0.89, science 0.78, and social studies 0.83.

Variables

The study had five dependent variables, including the four subject scores on the CBASE (i.e., English, mathematics, science, and social studies), and a composite score that combined all four subject tests. The independent variable was year in school (i.e., freshman, sophomore, junior, or senior). Because it was anticipated that sample mortality effects might bias upward the estimated effects of year in school on CBASE scores (e.g., seniors being a more academically select group than freshmen), controls for a number of important confounding variables were introduced. At the individual level controls were made for three precollege or demographic variables: academic ability (self-reported composite ACT score or SAT equivalent), sex, and race (i.e., African American, Latino, Asian American, Native American, and Caucasian). A series of preliminary analyses also included age, but its unique relationship with CBASE scores, and its contribution to the results, was so trivial that it was excluded. Also included were two additional individual-level control variables. The first was self-reported total credit hours taken in postsecondary education. It was judged important to include this variable to have at least some control for possible academic coursework taken at other institutions prior to the one at which the student was when he or she took the CBASE. The second was self-reported cumulative grade point average during college. Because grades appear to measure academic motivation and willingness to work hard as well as academic ability (Pascarella & Terenzini, 1991), it was thought important to control for cumulative grade point average as a reasonable proxy for a student's level of academic motivation.

Because evidence also suggests that the overall cognitive or academic ability of an institution's student-body can influence the climate of an institution (Astin, 1977, 1993; Pascarella & Terenzini, 1991), an estimate of aggregate student-body academic ability was also included as a control variable. This estimate was the mean self-reported ACT or SAT equivalent of the students enrolled in each institution in the study. Each individual student was then assigned the mean estimate of student-body academic aptitude at his or her institution.

Although it would have been preferable to have actual rather than self-reported measures of academic aptitude and college grades, there is clear evidence of a high correspondence on these measures. For example, Trusheim (1994) found correlations between self-reported scores and actual scores of 0.93 and 0.94 for SAT Verbal and SAT Quantitative scores, respectively. Similarly, reviewing nearly 40 years of research, Baird (1976) found that correlations between self-reported and actual grades ranged from 0.74 to 0.96. Furthermore, in our sample the self-reported measures appeared to have simple correlations with CBASE scores that were both reasonable in magnitude and consistent with previous evidence (e.g., Pascarella & Terenzini, 1991). For example, the self-reported ACT score had simple correlations with the CBASE composite score that ranged between 0.51 and 0.52, while the estimated institutional average ACT score had correlations with the CBASE composite score that ranged between 0.21 and 0.22. Self-reported college grades had correlations with the CBASE composite score that ranged between 0.30 and 0.32.

Final Sample

The study sample was comprised of students from 56 four-year postsecondary institutions with complete data on all variables. Only those institutions were included that had tested a minimum of 50 students, although because of the distribution, the effective minimum number of students was closer to 90. This yielded effective samples of 19,717 for the English test; 19,363 for the mathematics test; 19,461 for the science test; 19,848 for the social studies test and 18,418 in analyses of CBASE Composite. The sample was approximately 53% seniors, 17% juniors, 16% sophomores, and 14% freshmen.

Analyses

The primary analyses regressed each of the five CBASE scores on the set of six control variables (i.e., individual ACT score; race; sex; cumulative grade point average; total postsecondary credits taken; average institutional ACT score), and year in college. Race was represented by dummy variables representing African American, Asian American, Latino, and Native American, with Caucasian always coded 0. Year in college was represented by dummy variables for sophomore, junior, or senior, with freshman always coded 0. Thus, all resulting coefficients for year in college represented comparison with freshman students or those with the least exposure to postsecondary education. The unstandardized regression coefficient, or *b*, for sophomores, juniors, or seniors indicated the advantage (versus freshmen) on CBASE subject scores, net of the confounding variables. A separate set of regressions was also run to determine the unique variance increment (R^2 increase) attributable to year in college. These regressions were conducted in a setwise manner with all six of the control variables entered on step one, and the three dummy variables representing year in college entered on a subsequent step.

To determine the magnitude of the effects of year in college on CBASE scores, two "effect sizes" were estimated. The first, suggested by Hays (1994), divided the unstandardized regression weight (*b*) for sophomores, juniors, and seniors by the pooled sample standard deviation of the dependent variable. The resultant effect size represented that part of a standard deviation on specific CBASE subject scores that sophomores, juniors, or seniors were advantaged relative to their freshman counterparts. Computation of a second effect size was recommended by a reviewer. In this effect size estimate the *b* weight for year in college was divided by the square root of the mean square residual for the overall regression equation. Because this approach takes into account the reduction in the residual variance ($1-R^2$), it produces effect sizes that are larger than those produced when the *b* weight is divided by the pooled standard deviation of the dependent variable. One might consider our first approach to estimating effect size as providing a conservative estimate, whereas our second approach provides a more liberal estimate.

Before conducting the primary analyses of the study we sought to determine if the net effects of year in college differed in magnitude for students in different categories or different levels of the six control variables (i.e., individual ACT scores, sex, race, cumulative grade point, total postsecondary credits taken, average institutional ACT score). As indicated previously, these are termed "conditional effects." A set of cross-product terms was developed between the six control variables and the three dummy variables representing year in college. This set of cross-product terms was then added to the regression equations specified in the primary analyses (i.e., the six control variables plus the three dummy variables representing year in college). A significant increase in R^2 (explained variance) due to the set of cross-product terms indicated the presence of significant conditional effects for year in college (Pedhazur, 1982). If this condition was met, we then determined the significance of specific conditional effects in two steps. First, we determined if each specific conditional effect (for example, the set of three cross-product terms representing the conditional effect of individual ACT score x year in college) accounted for a significant R^2 increase over and above all main effects plus the set of cross-products representing all other conditional effects (Pedhazur, 1982). Second, we also determined if each specific conditional effect accounted for a significant increase in R^2 over and above the main effects.

Because of the extremely large sample sizes, a conservative alpha level was chosen. No effect was considered significant unless the null hypotheses could be rejected at $p < 0.001$.

Results

The overall tests for the presence of conditional effects were significant at $p < 0.001$ for all five dependent measures. The addition of the entire set of control variable x year in college cross-products to the main-effect equations were associated with R^2 increases of between 0.92% and 1.20%. However, only one conditional effect (sex x year in college) met both tests for the significance of an individual conditional effect. Controlling for all main effects and all other conditional effects, the three cross-product terms representing sex x year in college were associated with modest, but significant ($p < 0.001$) R^2 increases across all five dependent variables of between 0.65% and 0.90%. Similarly, with all main effects in the equation the three cross-product terms representing sex x year in college accounted for significant ($p < 0.001$) R^2 increases across all five dependent variables of between 0.70% and 0.95%. None of the other conditional effects met either of these criteria, even at $p < 0.05$. Such evidence suggests that the effects of college on CBASE scores did not differ in magnitude for students who differed in precollege academic ability, race, college grades, cumulative credits taken in postsecondary education, or the average academic ability of the students at the institution attended. On the other hand, the significant sex x year in college conditional effect suggests that the effects of college on

CBASE scores differ in magnitude for men and women. Consequently, we disaggregated the sample by sex and developed separate estimates for men and women.

Table 1 summarizes the estimates of net effects of each year in college (versus freshmen) on CBASE scores for men and women. The table shows a number of general trends. First, using the Composite score as a reasonable average of the four subject area scores, estimates of the overall net effects of college indicate that senior men have between 0.7 and 0.9 of a standard deviation advantage over freshman men, and senior women have between 0.5 and 0.6 of a standard deviation advantage over freshman women across all CBASE subject scores. These are typically considered moderate to large size effects (e.g., Bowen, 1977).

A second major trend supported in the data is that the vast majority of college impact on CBASE scores appears to take place during the first two years of college. This is evident from comparisons of the magnitude of the sophomore (versus freshman) effect sizes with the corresponding magnitudes of the senior (versus freshman) effect sizes. Again, using the Composite score as a reasonable average of the four subject area scores, it was found that the estimated sophomore effect for men, of between 0.708 and 0.868 of a standard deviation, was about 95% of the estimated senior effect for men, of between 0.742 and 0.909 of a standard deviation. Similarly, the corresponding average sophomore effect for women, of between 0.487 and 0.596 of a standard deviation, was about 98% of the estimated senior effect for women, of between 0.496 and 0.608 of a standard deviation.

Simply looking at the Composite score, however, does mask some interesting patterns of estimated effects attributable to year in college. For example, on both the English and social studies tests most of the estimated net effect of the college experience does occur by the sophomore year. However, students also appear to make modest additional progress in these areas through their senior year. On the English test, for example, the estimated sophomore (versus freshman) effect sizes for men and women ranged between 84% and 71% of the corresponding senior (versus freshman) effect sizes for men and women, respectively. Similarly, on the social studies test, the estimated sophomore effect sizes for women were about 75% of the corresponding senior (versus freshman) effect sizes. For both men and women on the English test, and for women on the social studies test, the unstandardized regression coefficients (*b*) for the estimated sophomore and senior effects were significantly different from each other at $p < 0.001$.

A different pattern for the net effects of the college experience emerged for the mathematics and science tests. On the mathematics test,

TABLE I

Estimated Effects of College on CBASE Scores for Men and Women

Year in College	English		Mathematics		Science		Social Studies		Composite	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Sophomore (vs. Freshman)										
b	38.230	24.700	33.500	27.400	44.130	31.450	36.070	20.160	39.840	27.390
Effect size estimate I[a]	0.612	0.396	0.488	0.400	0.603	0.430	0.559	0.313	0.708	0.487
Effect size estimate II[b]	0.718	0.464	0.586	0.474	0.716	0.510	0.663	0.371	0.868	0.596
Women's effect size as a percent of men's effect size	64.70		81.79		71.31		55.90		68.79	
Junior (vs. Freshman)										
b	38.410	24.880	27.740	21.490	43.780	34.310	35.470	19.240	39.170	27.150
Effect size estimate I[a]	0.616	0.399	0.404	0.313	0.598	0.469	0.550	0.298	0.696	0.483
Effect size estimate II[b]	0.721	0.467	0.485	0.376	0.709	0.556	0.652	0.353	0.853	0.591
Women's effect size as a percent of men's effect size	64.77		77.48		78.43		54.20		69.40	
Senior (vs. Freshman)										
b	45.630	35.070	36.390	15.280	43.500	30.510	38.430	26.850	41.740	27.900
Effect size estimate I[a]	0.731	0.561	0.531	0.223	0.595	0.417	0.596	0.416	0.742	0.496
Effect size estimate II[b]	0.857	0.659	0.631	0.266	0.705	0.495	0.708	0.494	0.909	0.608
Women's effect size as a percent of men's effect size	76.74		42.10		70.15		69.80		66.84	
Pooled Standard Deviation	62.43		68.58		73.15		64.51		56.24	
Square root of pooled mean square residual	53.21		57.18		61.67		54.39		45.92	
R^2 Total Model	0.247	0.285	0.272	0.299	0.255	0.289	0.217	0.261	0.354	0.395
R^2 Increase Due to Year in College	0.021	0.015	0.012	0.013	0.016	0.016	0.016	0.010	0.024	0.019

NOTES: Controls: ACT score; Race (African American, Asian American, Caucasian, Latino, Native American); College grades; Cumulative credits taken in postsecondary education; Average ACT score of the sample of students at the institution attended. All estimated effects (i.e., b, R^2 Total Model, R^2 Increase Due to Year in College) are statistically significant at $p < 0.001$.

[a] divided by sample pooled standard deviation.

[b] divided by square root of pooled mean square error from regression including main effects and cross-product terms for conditional effects.

the estimated sophomore (versus freshman) effect sizes for women were actually somewhat larger than the corresponding senior (versus freshman) effect sizes for women (about 1.8 times as large). Indeed, the *b* weight representing the sophomore effect for women (27.40) was significantly ($p < 0.001$) larger than the corresponding *b* weight representing the senior (versus freshman) effect for women (15.28). Thus, the estimates suggest that after making substantial progress in mathematics during the first two years of college (between 0.400 and 0.474 of a standard deviation) women tend to retrogress through their senior year. Despite this slipping backward, senior women are still functioning at about a quarter (0.223 to 0.266) of a standard deviation higher in mathematics than are freshman women. For men about 92% of the senior advantage over freshmen in mathematics comes by the end of the sophomore year. However, the estimated senior advantage over freshmen (0.531 to 0.631 of a standard deviation) is still larger, though not significantly so, than the advantage of sophomore men over their freshman counterparts (0.488 to 0.586 of a standard deviation).

On the science test, the progress made during the first two years of college by both men and women appears to stabilize and remain largely unchanged throughout the remainder of college. There was a general parity between the estimated sophomore (versus freshman) and senior (versus freshman) effects on science knowledge for both men and women. In both cases the b weight for the sophomore effect (44.13 for men and 31.45 for women) was not significantly different from the *b* weight for the senior effect (43.50 for men and 30.51 for women).

Finally, as suggested by the modest, but statistically significant sex x year of enrollment conditional effects across all five CBASE scores, there were non-chance differences in the effects of college for men and women. As Table 1 suggests, the impact of college on CBASE scores was substantially more pronounced for men than it was for women. Using the CBASE Composite score as a reasonable average across the four subject areas, the data from Table 1 clearly indicate that the estimated sophomore, junior, and senior (versus freshman) effect sizes for women were only about two-thirds (around 67%), of the corresponding effect sizes for men. If one considers the senior (versus freshman) net difference as an estimate of the overall effect of college, senior men had between 0.742 and 0.909 of a standard deviation advantage on the CBASE composite over their freshman counterparts. Senior women, on the other hand, had a net advantage over freshman women of between 0.496 and 0.608 of a standard deviation on the CBASE Composite score. The ratio of the senior (versus freshman) effect sizes for men and women on the both the science and social studies tests averaged about

0.70, or 70%; and the corresponding ratio of the senior effect sizes for men and women on the English test was about 0.77 or 77%. On the mathematics test, however, the advantage of senior women over their freshman counterparts was only about 42% of the corresponding advantage of senior men over freshman men. In light of such evidence it may be that men are making greater learning gains during college than women are, at least as measured by the CBASE.

Conclusions and Discussion

This study employed a cross-sectional design with statistical controls for important confounding influences to estimate the net impact of college on five standardized, objective measures of learning: English, mathematics, science, social studies, and a composite of all four tests. The findings suggest a number of general conclusions. First, estimates from samples of students attending 56 four-year colleges and universities suggest that, on average, across all subject areas, senior men have an advantage of between 0.74 and 0.91 of a standard deviation over freshman men. Senior women had an average advantage across all subject areas of between 0.50 and 0.61 of a standard deviation over their freshman counterparts. These are typically considered moderate to large effects, and they persisted net of statistical controls for individual ACT score, race, college grades, total credits taken in postsecondary education, and the average ACT score of students at the institution attended. For both men and women the estimated net effects of college varied among subject areas, from relatively smaller senior (versus freshman) advantages in mathematics (0.22 to 0.27 of a standard deviation for women and 0.56 to 0.63 of a standard deviation for men) to relatively larger corresponding advantages in English (0.56 to 0.66 of a standard deviation for women and 0.73 to 0.86 of a standard deviation for men).

A second major finding was that there were statistically significant differences by sex in the magnitude of the effects of college on CBASE scores. Our estimates suggest that the cognitive impacts of college may be substantially more pronounced for men than for women. Indeed, on average, across all subject areas, the net sophomore, junior, and senior (versus freshman) effects for women were only 69%, 69%, and 67% of the respective sophomore, junior, and senior (versus freshman) effects for men. In the senior year this ratio of female/male effects (expressed as a percentage) ranged from a low of about 42% in mathematics, to a high of about 77% in English.

Estimates using the CBASE would tend to support the contention that postsecondary education is a "gendered experience" that is not only manifestly different for women but may also be structured toward masculine rather than feminine intellectual orientations and learning styles (e.g., Baxter Magolda, 1992, 1998; Belenky, Clinchy, Goldberger, & Tarule, 1986; Hall & Sandler, 1982; Holland & Eisenhart, 1990; Levine & Cureton, 1998; Seymore, 1995; Whitt, 1992). To the extent this is the case, there may be sex-based classroom, organizational, or environmental differences woven into the fabric of postsecondary institutions that lead to women making smaller knowledge gains than men. For example, recent evidence from analyses of the National Study of Student Learning suggests that both classroom and nonclassroom actions that are perceived as unsupportive of women can have negative consequences for women's academic and cognitive growth during college (e.g., Pascarella, Whitt, Edison, Nora, Hagedorn, Yeager, & Terenzini, 1997; Whitt, Edison, Pascarella, Nora, & Terenzini, 1999). It is difficult to determine the influence of such factors from the CBASE data, however, because, unlike the National Study of Student Learning database, it does not contain extensive information regarding student classroom or nonclassroom experiences.

Of course there is an obvious alternative hypothesis to explain the gender differences we found in the effects of college on CBASE scores. Men and women tend to enroll in different majors (Jacobs, 1996; Pascarella & Terenzini, 1991), and attendant differences in coursework taken might well account for some, if not all, the differences we observed, particularly in mathematics and science. Although this possibility is indeed acknowledged, it is important to point out that the overall impact of college for women (i.e., the senior versus freshman effect) was significantly and substantially smaller than that for men across all the CBASE subject tests. Still, because the CBASE data do not contain information on major field of study or coursework taken, we cannot determine the extent to which such factors may have accounted for the gender differences we observed. This is a clear limitation of our study.

A third major finding from our estimates was that the vast preponderance of college impact on CBASE subject areas occurred during the first two years of postsecondary education. On average, across all subject tests, the net sophomore-freshman difference was about 95% to 98% of the net senior-freshman difference. There was modest continued improvement in English and social studies through the senior year; but on science there was little or no improvement, and on mathematics the net senior-freshman difference for women was actually smaller than the corresponding sophomore-freshman difference. Because the CBASE is designed to assess knowledge and intellectual skills learned as a result of general education courses, this overall trend is perhaps not surprising.

For students who do not major in quantitative fields most required mathematics courses may be completed in the first two years of college. Findings by Wolfle (1983) suggest that for those who do not major in quantitative fields, college tends to stabilize mathematical competencies at the levels at which students enter college. Our findings are in partial agreement with Wolfle's, but in the CBASE sample men and women had respective senior-year mathematics scores that were between 0.53 and 0.63 of a standard deviation, and between 0.22 and 0.27 of a standard deviation greater than those of their freshman counterparts.

A fourth major finding of the study was what was *not* found. Recall that tests for conditional effects indicated that, net of the control variables, including individual level ACT scores, the effect of college on CBASE scores did not differ for students who attended institutions differing in the average ACT scores of their student bodies. Put another way, the cognitive impact of college did not differ in magnitude for students attending institutions differing in academic selectivity. (Although our institutional sample did not include institutions of the very highest selectivity [such as Ivy League institutions] there was still a substantial amount of heterogeneity in average institutional ACT score. The estimated average institutional ACT range in the sample was about 18.50 to 28.10.) Such results are quite consistent with recent findings reported by Hagedorn, Pascarella, Edison, Braxton, Nora, & Terenzini (1999). They found that the average level of measured critical thinking of the students at an institution had no significant effect on growth in individual student critical thinking after three years of college. Our estimates are also consistent with the conclusion of Pascarella & Terenzini (1991) in their review of 20 years of research on the impact of college on students; namely, institutional selectivity, in and of itself, may have only trivial and inconsistent net impacts on the cognitive development or learning gains of individual students. Institutions that have relatively stringent admissions standards may not always provide the most impactful intellectual environment. Programmatic efforts and the quality of teaching may be more salient determinants of an institution's impact on student learning and cognitive growth than the average academic selectivity of the student body.

References

Anaya, G. (1999a). *Accuracy of student-reported test scores: Are they suited for college impact assessment?* Unpublished manuscript, Indiana University, Bloomington, IN.

Anaya, G. (1999b). College impact on student learning: Comparing the use of self-reported gains, standardized test scores, and college grades. *Research in Higher Education, 40,* 499–526.

American College Testing Program (1993). *CAAP user norms: Fall, 1993.* Iowa City, IA: Author.

American Psychological Association. (1985). *Standards for educational and psychological testing.* Washington, DC: Author.

Astin, A. (1977). *Four critical years.* San Francisco: Jossey-Bass.

Astin, A. (1993). *What matters in college.* San Francisco: Jossey-Bass.

Baird, L. (1976). *Using self-reports to predict student performance.* Research Monograph No. 7. New York: College Entrance Examination Board.

Bauer, K. (1996, May). *Year two of the longitudinal study of university undergraduates: Changes occurring from the freshman to sophomore year.* Paper presented at the meeting of the Association for Institutional Research, Albuquerque, NM.

Bauer, K. (1998, May). *Academic and social development of undergraduate students: Summary of findings for the UD longitudinal study, Fall 1993–Spring 1997.* Paper presented at the meeting of the Association for Institutional Research, Minneapolis, MN.

Baxter Magolda, M. (1990). Gender differences in epistemological development. *Journal of College Student Development, 31,* 555–561.

Baxter Magolda, M. (1991, April). *A gender inclusive model of epistemological development.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Baxter Magolda, M. (1992). *Knowing and reasoning in college: Gender-related patterns in students' intellectual development.* San Francisco: Jossey-Bass.

Baxter Magolda, M. (1998). Measuring gender differences in intellectual development: A comparison of assessment methods. *Journal of College Student Development, 29,* 528–537.

Belenky, M., Clinchy, B., Goldberger, N., & Tarule, J. (1986). *Women's ways of knowing: The development of self, voice, and mind.* New York: Basic Books.

Bowen, H. (1977). *Investment in learning.* San Francisco: Jossey-Bass.

Bradburn, N., & Sudman, S. (1988). *Polls and surveys.* San Francisco: Jossey-Bass.

Dollar, R. (1991). College influence on analytical thinking and communication skills of part-time versus full-time students. *College Student Journal, 25,* 273–279.

Durham, R., Hays, J., & Martinez, R. (1994). Socio-cognitive development among Chicano and Anglo American college students. *Journal of College Student Development, 35,* 178–182.

Evans, D. (1989). Effects of a religious-oriented, conservative, homogeneous college education on reflective judgment. *Dissertation Abstracts International, 49,* 3306-A.

Feldman, M. (1994). A strategy for using student perceptions in the assessment of general education. *Journal of General Education, 43,* 151–167.

Flowers, L., & Pascarella, E. (1999). Cognitive effects of college racial composition on African American students after three years of college. *Journal of College Student Development, 40,* 669–677.

Gardner, L. (1994). *Redesigning higher education: Producing dramatic gains in student learning.* Washington, DC: Clearinghouse on Higher Education, The George Washington University.

Giancarlo, C., & Facione, P. (1997). *A longitudinal study of Santa Clara undergraduate students' dispositions toward critical thinking.* Unpublished manuscript, Santa Clara University, Santa Clara, CA.

Hagedorn, L., Pascarella, E., Edison, M., Braxton, J., Nora, A., & Terenzini, P. (1999). Does institutional context influence the development of critical thinking? A research note. *Review of Higher Education, 22,* 247–263.

Hall, R., & Sandler, B. (1982). *The campus climate: A chilly one for women?* Report of the Project on the Status and Education of Women. Washington, DC: Association of American Colleges.

Hannah, L., & Michaelis, J. (1977). *A framework for instructional objectives: A guide to systematic planning and evaluation.* Reading, MA: Addison-Wesley.

Hart, J., Rickards, W., & Mentowski, M. (1995, April). *Epistemological development during and after college: Longitudinal growth on the Perry scheme.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hays, W. (1994). *Statistics (5th ed.).* Fort Worth, TX: Harcourt Brace College Publications.

Hill, D. (1995). Critical thinking and its relation to academic, personal, and moral development in the college years. *Dissertation Abstracts International, 56,* 4603-B.

Holland, D., & Eisenhart, M. (1990). *Educated in romance: Women, achievement, and college culture.* Chicago: University of Chicago Press.

Jacobs, J. (1996). Gender inequity and higher education. *Annual Review of Sociology, 22,* 153–185.

Jehng, J., Johnson, S., & Anderson, R. (1993). Schooling and students' epistemological beliefs about learning. *Contemporary Educational Psychology, 18,* 23–35.

King, P., & Kitchener, K. (1994). *Developing reflective judgment.* San Francisco: Jossey-Bass.

Kitchener, K. (1992). *The development of critical reasoning: Does a college education make a difference?* (University Lecture). Denver, CO: University of Denver.

Kube, B., & Thorndike, R. (1991). *Cognitive development during college: A longitudinal study measured on the Perry scheme.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Kuh, G. (1993). In their own words: What students learn outside the classroom. *American Educational Research Journal, 30,* 277–304.

Kuh, G., Douglas, K., Lund, J., & Ramin-Gyurnek, J. (1994). *Student learning outside the classroom: Transcending artificial boundaries.* ASHE-ERIC Higher Education Report No. 8. Washington, DC: George Washington University, School of Education and Human Development.

Levine, A., & Cureton, J. (1998). *When hope and fear collide.* San Francisco: Jossey-Bass.

Lincoln, C. (1991). The relation between college orientation and student effort: An exploratory study. *Dissertation Abstracts International, 51,* 4034–A.

May, D. (1990). *Student development: A longitudinal investigation of the relationship between student effort and change in intellectual development.* Unpublished doctoral dissertation, Memphis State University, Memphis, TN.

McDonough, M. (1997). An assessment of critical thinking at the community college level. *Dissertation Abstracts International, 58,* 2516-A.

Myerson, J., Rank, M., Raines, F., & Schnitzler, M. (1998). Race and general cognitive ability: The myth of diminishing returns to education. *Psychological Science, 9,* 139–142.

Osterlind, S. (1997). *A national review of scholastic achievement in general education: How are we doing and why should we care?* ASHE-ERIC Higher Education Report Volume 25, No. 8. Washington, DC: The George Washington University, Graduate School of Education and Human Development.

Osterlind, S., Robinson, R., & Nickens, N. (1997). Relationship between collegians' perceived and congeneric tested achievement in general education. *Journal of College Student Development, 38,* 255–265.

Pascarella, E. (1989). The development of critical thinking: Does college make a difference? *Journal of College Student Development, 30,* 19–26.

Pascarella, E., Bohr, L., Nora, A., & Terenzini, P. (1996). Is differential exposure to college linked to the development of critical thinking? *Research in Higher Education, 37,* 159–174.

Pascarella, E., & Terenzini, P. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Pascarella, E., Whitt, E., Edison, M., Nora, A., Hagedorn, L., Yeager, P., & Terenzini, P. (1997). Women's perceptions of a "chilly climate" and their cognitive outcomes during the first year of college. *Journal of College Student Development, 38,* 109–124.

Pearson, F., & Rodgers, R. (1998). Cognitive and identity development: Gender effects. *Initiatives, 58,* 17–33.

Pedhazur, E. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). New York: Holt, Rinehart & Winston.

Pike, G. (1991). Assessment measures. *Assessment Update: Progress, Trends, and Practices in Higher Education, 3*(1), 6–7.

Pike, G. (1992). The components of construct validity: A comparison of two measures of general education outcomes. *Journal of General Education, 41,* 130–160.

Pike, G. (1995). The relationship between self-reports of college experiences and achievement test scores. *Research in Higher Education, 36,* 1–21.

Pike, G. (1996). Limitations of using students' self-reports of academic achievement measures. *Research in Higher Education, 37,* 89–114.

Rykiel, J. (1995). The community college experience: Is there an effect on critical thinking and moral reasoning? *Dissertation Abstracts International, 56,* 3824-A.

Saucier, B. (1995). Critical thinking skills of baccalaureate nursing students. *Journal of Professional Nursing, 11,* 351–357.

Seymore, E. (1995). The loss of women from science, mathematics, and engineering undergraduate majors: An exploratory account. *Science Education, 79,* 437–473.

Smith-Sanders, C., & Twale, D. (1997, March). *Impacts of number and type of core curriculum hours on critical thinking.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Smith-Sanders, C., & Twale, D. (1998, March). *Further study of the impact of number and type of core curriculum hours on critical thinking.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Stage, F., Muller, P., Kinzie, J., & Simmons, A. (1998). *Creating learning centered classrooms: What does learning theory have to say?* (ASHE-ERIC Higher Education Report, No. 4.) Association for the Study of Higher Education.

Steel, J., & Nichols, P. (1992, November). *Linking COMP objective test scores with student perceptions of growth.* Paper presented at the annual meeting of the American Evaluation Association, Seattle, WA.

Tan, D. (1995). Do students accomplish what they expect out of college? *College Student Journal, 29,* 449–454.

Thorndike, R., & Andrieu-Parker, J. (1992, April). *Growth in knowledge: A two-year longitudinal study of changes in scores on the College Basic Academic Subjects Examination.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Thorndike, R., Andrieu-Parker, J., & Kube, B. (1990, April). *Impact of college experiences on cognitive development.* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Trusheim, D. (1994). How valid is self-reported financial aid information? *Research in Higher Education, 35,* 335–348.

Whitmire, E., & Lawrence, J. (1996, November). *Undergraduate students' development of critical thinking skills: An institutional and disciplinary analysis and comparison with academic library use and other measures.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Memphis, TN.

Whitt, E. (1992, November). *"Taking women seriously": Lessons for coeducational institutions from women's colleges.* Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Minneapolis, MN.

Whitt, E., Edison, M., Pascarella, E., Nora, A., & Terenzini, P. (1999). Women's perceptions of a "chilly climate" and cognitive outcomes in college: Additional evidence. *Journal of College Student Development, 40,* 163–177.

Williams, R. (1996). Assessing students' gains from the college experience at East Tennessee State University. *Dissertation Abstracts International, 57,* 970–A.

Wolfle, L. (1983). Effects of higher education on achievement of blacks and whites. *Research in Higher Education, 19,* 3–9.

Wood, P. (1997). A secondary analysis of claims regarding the Reflective Judgment Interview: Internal consistency, sequentiality, and intro-divisional differences in ill-structured problem solving. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 12, pp. 243–312). New York: Agathon.

Zhang, Z., & RiCharde, R. (1999). Intellectual and metacognitive development of male college students: A repeated measures approach. *Journal of College Student Development, 40,* 721–738.

Laura W. Perna

The Relationship Between Family Responsibilities and Employment Status Among College and University Faculty

Although the participation of mothers in the labor force is viewed more favorably now than in the past, a substantial proportion of American workers continue to believe that women should focus their efforts on the home (Bond, Galinsky, & Swanberg, 1997). For example, surveys by the Families and Work Institute revealed that 41% of employees nationwide agreed in 1997 that men should be the breadwinner and women should care for the home and children, down from 64% in 1977 (Bond, Galinsky, & Swanberg, 1997).

Research suggests that college and university faculty also perceive tension between work and family roles (Cole & Zuckerman, 1987; Finkel, Olswang, & She, 1994; Marshall & Jones, 1990; Sorcinelli & Near, 1989). For instance, from their exploratory study of 12 women and minorities who had made choices about entering academia, Bronstein, Rothblum, and Solomon (1993) concluded that the concentration of women in nontenure-track and part-time positions was due, in part, to the conflict between career and family demands. Through interviews, Cole and Zuckerman (1987) found that even the youngest women scientists in their sample, women who had received their doctorates during the 1970s, encountered individuals who viewed marriage and motherhood to be incompatible with a scientific career. Although Marshall and Jones (1990) found that the timing of childbearing was unrelated to salaries and academic rank among female higher education deans, ad-

The author is grateful for the comments and suggestions of Jeffrey Milem and three anonymous reviewers.

Laura W. Perna is assistant professor, Department of Education Policy and Leadership, University of Maryland at College Park.

The Journal of Higher Education, Vol. 72, No. 5 (September/October 2001)
Copyright © 2001 by The Ohio State University

ministrators, and counselors, they also found that about two-thirds of their sample believed that childbearing had negatively affected their careers, particularly in terms of their professional advancement and mobility. A survey of tenured and tenure-track faculty at one university showed that the majority (70%) believed that taking leave after the birth of a child would be detrimental to their careers (Finkel et al., 1994).

The lower representation of married women than single women, and women with children than childless women, among the nation's college and university faculty may also suggest the difficulties associated with fulfilling both family and career responsibilities. Analyses of the 1993 National Study of Postsecondary Faculty (NSOPF:93) reveal that women represent a smaller share of married than never married faculty (34% versus 52%) and a smaller share of faculty with children than childless faculty (31% versus 54%). Some research (e.g., Cooney & Uhlenberg, 1989) suggests that these patterns are similar to those for other professional women. For example, using 1980 census data, Cooney and Uhlenberg (1989) showed that White women lawyers, physicians, and postsecondary teachers were substantially less likely than White women of the same age in the general population to be married and have children. Among White women between the ages of 35 and 39 with at least ten years of marriage, a substantially higher share of postsecondary teachers than of physicians or lawyers were childless (Cooney & Uhlenberg, 1989).

Although the challenges associated with balancing work and family roles may not be unique to faculty, these data raise important questions not only about the extent to which marriage and motherhood may limit access to a faculty career in general but also about the specific types of academic positions that are available to married women and women with children. Of particular concern is the extent to which married women and women with children may be concentrated in lower-status faculty positions. Although the representation of women among college and university faculty has increased since the mid-1970s, the greatest growth has been among part-time and nontenure-track appointments (Chronister, Gansneder, Harper, & Baldwin, 1997; Lomperis, 1990). While the number of nontenured but tenure-track full-time faculty increased between 1976 and 1993 at a faster rate for women than men (22% increase for women versus 24% decline for men), the greatest growth has been among non-tenure track appointments. Between 1976 and 1993, the number of non-tenure track full-time faculty increased by 142% for women and 54% for men (Chronister et al., 1997). In part because of these differential growth rates, analyses of the NSOPF:93 show that women represented a higher share of full-time, nontenure-track faculty than of full-time, tenure-track faculty in fall 1992 (52% versus 43%).

Although anecdotal evidence (e.g., Flynn, Flynn, Grimm, & Lockhart, 1986; Wilson, 1998) suggests that not all nontenure-track faculty are dissatisfied with their status, many nontenure-track faculty may be considered to be marginal "in the sense that they hope for full integration into academe" (Bowen & Schuster, 1986, p. 65), and because they represent a lower rung on the hierarchy of academic labor markets (Youn, 1992). According to Youn (1992), the existence of hierarchies within the academic labor market contributes to various forms of segmentation including segmentation by job status (e.g., full-time or part-time). Movement from one job status segment to another (e.g., from part-time to full-time or from nontenure-track to tenure-track) is restricted, just as is movement from one academic discipline to another (e.g., from mathematics to English). Competition among faculty in different segments is limited, thereby permitting inequities among faculty across segments (Youn, 1992).

This study uses data from the NSOPF:93 to explore the extent to which the higher observed representation of women among nontenure-track faculty (i.e., the lower status positions) is related to family responsibilities after taking into account other variables that are expected to be related to employment status. Sex differences in the relationship between family responsibilities and employment status are examined and implications of the findings are discussed.

Theoretical Framework

Although little is known about the relationship between family responsibilities and employment status among college and university faculty, researchers have explored the relationship between family responsibilities and such outcomes as research productivity and salaries. From her comprehensive review and synthesis of prior research, Creamer (1998) concluded that most research shows no relationship between marital status and publishing productivity for women. In fact, some evidence suggests that married faculty are more productive than other faculty after controlling for other differences (Bellas & Toutkoushian, 1999). Some researchers (e.g., Astin & Bayer, 1979; Astin & Davis, 1985) have shown that married women, and others (e.g., Bellas, 1992; Hamovitch & Morgenstern, 1977) have shown that married men are more productive than their single counterparts of the same sex. The extent to which the occupation of the spouse is related to scholarly productivity is equivocal. Using a national sample of faculty employed at colleges and universities in 1989, Astin and Milem (1997) found that, after controlling for differences in background and job-related characteristics, having an academic spouse was associated with higher levels of research productivity for women but lower levels of research productivity for men. In contrast, using a sample of faculty employed in the state of Illinois in 1993, Bellas (1997b) found that having an academic partner was unrelated to research productivity after taking into account differences in other variables.

With regard to the relationship between parental responsibilities and research productivity, Creamer (1998) also concluded from her review of prior research that the relationship is ambiguous. She found no significant relationship between having children and publishing productivity in five of the ten studies reviewed, a positive relationship in three of the ten, and a negative relationship in two of the ten. Using a subsample from the NSOPF:93, Bellas and Toutkoushian (1999) showed that full-time faculty with dependents had higher levels of research productivity than full-time faculty without dependents after controlling for differences in sex, race, education, experience, academic field, institutional type, and allocation of time.

In terms of the relationship between family responsibilities and faculty salaries, Johnson and Stafford (1974), using data from the 1970 National Science Foundation Register, showed that labor force participation among women faculty was influenced by marital status, husband's earnings, and number of children and that time out of the labor force for child bearing and child rearing was negatively related to earnings. Whereas Barbezat (1988), after controlling for other variables, found marital status to be unrelated to the salaries of women and men faculty in both 1968 and 1977, others have shown that married men faculty received higher salaries than their single male counterparts in both 1984 (Bellas, 1992) and 1992 (Toutkoushian, 1998). Some research (Bellas, 1992; Astin & Milem, 1997) suggests that the employment status of the spouse or partner also matters. Bellas (1992) found that men faculty with nonemployed wives averaged higher salaries than men faculty with employed wives even after controlling for education, experience, productivity, rank, institutional characteristics, and academic field. Astin and Milem (1997) found that, after controlling for differences in background and job-related characteristics, having an academic spouse was associated with higher salaries among women faculty but lower salaries among men faculty. Some research (Barbezat, 1988) indicates that men faculty, but not women faculty, with children receive higher salaries.

In one of the few examinations of the relationship between family responsibilities and employment status, Ferber and Hoffman (1997) found that neither the probability of being employed at a research or doctoral university nor the probability of holding the highest rank of full professor were related to such measures of household responsibilities as geographic distance from the current partner, number of years spent with partners, level of education of partners, number of years partners employed at the same institution, number of children, and number of years children spent in the household among women faculty employed at colleges and universities in the state of Illinois in 1993.

Prior research has drawn upon two perspectives to examine gender equity issues in academic employment: human capital and structural. According to human capital theory, an individual's status and rewards in the academic labor market are determined primarily by his or her productivity. Productivity is expected to be determined by the investments that individuals make in themselves, particularly the quantity and quality of their education and the amount of their on-the-job training, as well as their geographic mobility, their motivation and intensity of work, and their emotional and physical health (Becker, 1962, 1993).

Some economists have argued that family responsibilities influence investment in human capital, continuity of labor force participation, types of employment sought, and level of commitment to the job (Becker, 1985; Polachek, 1977). Even today family responsibilities continue to be borne primarily by women (Bond et al., 1997). For example, a 1997 national survey of workers in a variety of occupations showed that married employed women spent more time than married employed men caring for their children and engaging in household chores on both workdays and non-workdays (Bond et al., 1997).

With regard to the accumulation of human capital, an individual who is out of the labor force because of family responsibilities is not acquiring additional on-the-job experience and may even be losing some previously acquired job skills (Becker, 1993). Korenman and Neumark (1991) noted that most prior research has concluded that children reduce wages indirectly by reducing labor force participation and the "accumulation of human capital" (e.g., experience) rather than by directly reducing productivity. Arguing that prior estimates may be biased, however, Korenman and Neumark (1991) concluded that children may directly reduce women's wages and that, because of this reduction in wages, women reduce their participation in the labor market. Regardless of the direction of causality, women with children appear to average lower levels of experience.

Family responsibilities may also reduce geographic mobility. Research has shown that women are less mobile than men (Marwell, Rosenfeld, & Spilerman, 1979; Rosenfeld & Jones, 1987). For example, women have been found to be more likely than men to remain in the geographic area where they attended graduate school and to be concentrated in larger urban areas where, presumably, the probability of both partners finding satisfactory employment is higher (Marwell et al., 1979; Rosenfeld & Jones, 1987). The advantages of geographic mobility are evidenced by research showing that faculty who are more mobile receive higher salaries (Astin & Bayer, 1979; Smart & McLaughlin, 1978; Kasten, 1984), are more likely to hold tenure track positions (Rosenfeld & Jones, 1987), and tend to hold higher academic rank (Marwell et al., 1979) than other faculty.

Family responsibilities may also be related to the level of motivation and intensity of work. Human capital theorists (e.g., Becker, 1985) predict that, compared with men and single women, married women pursue less demanding jobs, such as part-time and nontenure-track positions, because household responsibilities require more effort than leisure and other nonmarket activities and, consequently, they have less energy available for market work. In other words, differences in household responsibilities are expected to be associated with differences in motivation and intensity of work and are expected to lead to occupational segregation by sex (Becker, 1985).

Marriage and parenting responsibilities may also influence emotional and physical health. As Tack and Patitu (1992) observed from their comprehensive review of the predictors of job satisfaction among women and minority faculty, "life-style stressors" (e.g., child care, parent care, physical and mental health) likely have a stronger impact on women than men because of societal expectations about the priority women should place on their families. Some (Austin & Pilat, 1990) have speculated that women may also feel greater stress and pressure than men in their attempts to balance work and family responsibilities because of the physical demands of pregnancy, childbirth, and early parenthood.

Household responsibilities and children's problems have been shown to be more important sources of stress for women faculty than for men faculty (Dey, 1994). Among full-time tenure-track faculty at one university, 28% more women than men reported experiencing at least one of the following conflicts between work and child care: avoiding overnight conferences because of child care demands, bringing a child to the university, delaying promotion or tenure because of child care responsibilities, or being unavailable to attend a function at the child's school because of work demands (Riemenschnieder & Harper, 1990). About two-thirds of women faculty, but only one-third of men faculty, reported feeling overwhelmed trying to meet both child care and employment demands (Riemenschnieder & Harper, 1990). Findings from another single institution study suggest that women faculty are as involved with their work as men, but that women are more likely than men to forego leisure activities in order to satisfy work-related demands (Sorcinelli & Near, 1989).

Despite the popularity of human capital theory for explaining differences in labor market experiences, some economists and sociologists have noted the limitations of this theory (DeYoung, 1989; Dreijmanis, 1991). Such critics argue that "focusing on the supply of human skills to explain economic inequality and lack of productivity is a theoretical mistake" (DeYoung, 1989, p. 155). Among the limitations is the inability of human capital theory to adequately explain the lower returns to educational investments among women and minorities (DeYoung, 1989). Based on her examination of the relationship between time out of the labor force and occupational choice, England (1982) concluded that "human capital theory has not generated an explanation of occupational sex segregation that fits the evidence" (p. 358). Contrary to the predictions of human capital theory, England's analyses of data from the National Longitudinal Survey showed that women with more continuous employment histories or plans were no less likely than other women to work in predominantly female occupations, the presumably lower-status occupations. In a test of Becker's (1985) assertion that women seek less demanding jobs and devote less effort to their jobs than men, Bielby and Bielby (1988) found that, after controlling for household responsibilities, earnings, and job responsibilities, women actually allocate more effort to their work than men.

Social scientists interested in the issues of social inequality and poverty have responded to the inadequacies of human capital theory by developing structural or institutional approaches to understanding labor markets (Youn, 1988). This perspective of labor market experiences generally emphasizes the effects of the attributes of the organizations with which individuals are connected, particularly the influence of the characteristics of the colleges and universities in which they were trained and in which they work, including these institutions' financial resources, tenure system, academic governance, and collective bargaining agreements.

Structural models posit that sex differences in the labor market experiences of faculty are attributable to the segregation of women in the types of institutions, academic fields, and work roles that have lower prestige and value (Smart, 1991). For example, Sorenson (1989) found that 20% of the national male-female wage difference in 1983 for all occupations, not just for faculty or higher education positions, was attributable to occupational segregation by sex after controlling for personal characteristics (e.g., tenure on the job, educational attainment, and full- or part-time status), characteristics of the occupation (e.g., education and training required to perform the job and working conditions), and attributes of the firm (e.g., geographic region, union status, size of firm, and major industry category). In higher education, the average salaries of faculty in institutions and disciplines with higher proportions of women have also been found to be lower than the average salaries of faculty in institutions and disciplines with smaller proportions of women (Barbezat, 1988; Bellas, 1994, 1997a; Smart, 1991).

Research Method

Although researchers have examined the effects of marital and parental status on research productivity and salaries, little is known about the relationship between family responsibilities and the employment status of faculty. Given the recent growth in nontenure-track positions, such research is particularly timely. Because some (e.g., Smart, 1991) have concluded that substantial research supports the appropriateness of both human capital and structural approaches to academic labor markets, this study draws upon both perspectives to explore the following research questions:

1. Are family responsibilities related to the employment status of women and men junior faculty after controlling for differences in human capital and structural characteristics?
2. To what extent are junior faculty satisfied holding nontenure-track positions?

Sample

Data from the NSOPF:93 are used to address the research questions. Sponsored by the U.S. Department of Education's National Center for Education Statistics, the NSOPF:93 is designed to provide a national profile of faculty, particularly with regard to their professional backgrounds, responsibilities, workloads, salaries, benefits, and attitudes. In the first stage of the two-stage sample selection, 974 public and private nonproprietary higher education institutions were selected and 817 agreed to participate. In the second stage, approximately 42 faculty and instructional staff were selected from each participating institution. A total of 25,780 questionnaires were returned by the 31,354 faculty and instructional staff who were sampled. For additional details on the survey methodology, refer to Kirshstein, Matheson, and Jing (1997).

The sample used in this research is limited to junior faculty whose primary responsibility is teaching. Junior faculty are defined as individuals with faculty status who hold tenure-track or nontenure-track positions. Tenured faculty and faculty who work at colleges and universities that do not have a tenure system are excluded from the sample.

The NSOPF:93 weight (WEIGHT) is appropriate for approximating the population of college and university faculty from the sample. To minimize the influence of large sample sizes and correct for the non-simple random sample design on standard errors, each case is weighted by the NSOPF:93 weight divided by the average weight for the sample (average weight = 37.72). The adjusted weighted sample includes 6,505 cases and represents 245,382 junior faculty nationwide.

Variables

The hypothesized predictors of employment status include measures of family responsibilities, human capital, and structural characteristics. In addition to sex, four racial/ethnic groups are considered: Black, Hispanic, Asian, and White. White is the reference group. Family responsibilities are measured by marital status and parental status. Marital status is measured by three dichotomous variables: married (married or living with someone); previously married (separated, divorced, or widowed); and never married (reference category). Parental status is measured by whether the individual has at least one child (yes or no).

Human capital is accumulated via educational attainment, on-the-job training, experience, and mobility (Becker, 1962). The level of investment in formal education is measured by a three-level categorical variable: whether an individual holds less than a doctoral degree; whether an individual holds a doctoral degree from a non-Research I university; and whether an individual holds a doctoral degree from a Research I university (reference category). On-the-job training is measured by whether the individual held a research assistantship and/or a teaching assistantship during graduate school. Experience is measured by the number of years since receiving the highest degree and the number of years in the current position (correlation = 0.405). Whether the individual is in his or her first or only job since having earned the highest degree (yes or no) is the best available proxy for mobility.

Structural attributes describe the type of institution and academic discipline in which a faculty member works. A categorical variable reflecting institutional Carnegie classification measures institutional resources as well as occupational segregation by institutional type. The categories are Research I (reference category), other Doctoral, Comprehensive I, other four-year, other (e.g., specialized), and two-year. The existence of collective bargaining agreements is measured by whether the faculty at the institution are unionized (yes or no). A measure of the tenure system, another structural attribute, is not necessary because only junior (i.e., nontenured) faculty working at institutions with a tenure system are included in the analyses. With regard to academic field, 12 categories are included to reflect substantive similarities as well as similarities in the representation of women among junior faculty in the field. The categories are nursing and other health (76% women), English and foreign languages (67%), education (66%), fine arts (49%), psychology, sociology, and other social sciences (48%), biology (38%), mathematics and computer science (33%), business (31%), history, philosophy, law, economics, and political science (31%), first-professional health (30%), engineering and physical science (15%), and other field (40% women). The reference category is engineering and physical science.

Analyses

The dependent variable, employment status, has four categories: employed full-time on a tenure track; employed part-time on a tenure track; employed full-time, not on a tenure track; and employed part-time, not on a tenure track. Because of the categorical nature of the dependent variable, a multinomial logit model, a special case of the general log-linear model, is used to examine the relationship between family responsibilities and employment status after controlling for human capital and structural characteristics. Part-time, tenure-track faculty are excluded from the multinomial logit analyses for two reasons. First, because the number of cases with part-time, tenure-track employment status is small (adjusted weighted sample size = 131, 2% of all junior faculty), including a separate category for part-time, tenure-track employment would result in problems with zero cells and unstable estimates of coefficients and standard errors (Menard, 1995).

Second, combining part-time, tenure-track with one of the other groups is inappropriate because the descriptive analyses show that part-time, tenure-track faculty are different from other faculty in several respects. Table 1 shows that part-time employment appears to be more common at public two-year institutions regardless of tenure status, with about 44% of part-time, tenure-track and part-time, nontenure-track faculty employed at public two-year institutions. Part-time, tenure-track faculty are relatively less common at Research I universities, with only 3% of part-time, tenure-track faculty but 12% of all junior faculty working at Research I universities. Part-time, tenure track faculty have generally received their highest degrees more recently and have worked fewer years in their current positions than other junior faculty. Though marital status appears to be unrelated to employment status, a higher share of part-time, tenure-track faculty than of all junior faculty have at least one child (78% versus 62%).

Therefore, only three employment categories are considered in the multinomial logit model: (1) employed full-time, on a tenure track; (2) employed full-time, not on a tenure track; and (3) employed part-time, not on a tenure track. Two contrasts are possible with three outcome categories. Full-time, nontenure-track employment and part-time, nontenure-track employment are simultaneously contrasted to full-time, tenure-track employment.

Multinomial logit models estimate the log-odds of one outcome oc-

TABLE 1

Selected Characteristics of Women and Men Junior Faculty by Employment Status: Fall 1992

Characteristic	Total	Full-time Tenure Track	Part-time Tenure Track	Full-time Nontenure Track	Part-time Nontenure Track
Weighted sample	245,382	4,928	87,592	35,442	117,420
Adjusted weighted *n*	6,505	131	2,322	940	3,113
Distribution	100.0%	35.7%	2.0%	14.4%	47.9%
INSTITUTIONAL TYPE	100%	100%	100%	100%	100%
Research I	12.0%	11.6%	3.1%	17.1%	11.1%
Other doctoral	14.7%	19.4%	13.2%	19.0%	10.0%
Comprehensive I	25.2%	30.1%	20.2%	28.4%	20.7%
Other 4-year	12.3%	15.3%	10.1%	14.5%	9.4%
Public 2-year	29.7%	18.4%	45.0%	12.1%	42.8%
UNIONIZED INSTITUTION	44.8%	41.1%	51.1%	41.5%	48.3%
WOMEN	46.0%	43.0%	46.6%	52.3%	46.4%
MARRIED					
Total	72.9%	72.3%	75.6%	71.0%	73.8%
Men	78.1%	80.0%	77.1%	74.2%	77.8%
Women	66.6%	62.0%	73.8%	67.9%	69.1%
PREVIOUSLY MARRIED					
Total	12.8%	12.3%	9.2%	12.1%	13.5%
Men	9.3%	7.9%	2.9%	10.7%	10.4%
Women	16.9%	18.1%	18.0%	13.6%	17.2%
AT LEAST 1 CHILD					
Total	62.1%	62.6%	77.7%	58.2%	62.2%
Men	69.7%	71.0%	82.9%	62.6%	70.1%
Women	53.1%	51.5%	70.5%	54.2%	53.1%
LESS THAN DOCTORAL DEGREE					
Total	66.2%	40.3%	81.3%	71.4%	83.5%
Men	61.0%	34.6%	76.5%	69.2%	79.4%
Women	72.1%	47.8%	86.7%	73.4%	88.0%
NO. YEARS SINCE HIGHEST DEGREE					
Total	12.1	13.0	8.7	11.8	14.9
Men	13.0	13.8	8.9	13.6	16.1
Women	11.1	12.1	8.3	10.1	13.4
NO. YEARS IN CURRENT POSITION					
Total	5.5	7.7	4.2	6.0	6.2
Men	6.0	8.8	4.4	6.3	7.1
Women	4.9	6.3	4.0	5.7	5.2

SOURCE: Analyses of 1993 National Study of Postsecondary Faculty (NSOPF:93).

curring relative to the baseline category (i.e., full-time, tenure track employment). If the baseline category is *J*, the model for the i^{th} category (e.g., full-time, nontenure-track employment) is:

$$\mathrm{Log}(P_i/P_j) = B_{i0} + B_{i1}X_1 + B_{i2}X_2 + \ldots + B_{ip}X_p$$

The logistic coefficients that result from this equation may be interpreted as the change in log odds associated with a one-unit change in the independent variable. The interpretation of the multinomial logit coefficients is facilitated by the use of odds-ratios, as described by the following equation:

$$P_i / P_j = e^{B_{i0} + B_{i1}X_1 + \ldots + B_{ip}X_p} = e^{B_{i0}} e^{B_{i1}X_1} \ldots e^{B_{ip}X_p}$$

The odds-ratio represents the change in the odds of choosing a particular employment status relative to the reference employment status (full-time tenure-track) that is associated with a one-unit change in a particular independent variable. An odds-ratio greater than one represents an increase in the likelihood of part-time or full-time, nontenure-track employment relative to full-time, tenure-track employment, whereas an odds-ratio less than one represents a decrease in the likelihood of part-time or full-time, nontenure-track employment.

The two continuous variables, number of years since receiving the highest degree and number of years in the current position, are entered into the model as covariates. The test of whether a coefficient is different from zero is based on the Wald statistic, which is calculated as the coefficient divided by its standard error, squared. Goodness-of-fit is reflected by the change in –2 log likelihood. A pseudo-R^2 is reported to provide an indication of the strength of the relationship between the outcome variable and the independent variables.

Limitations

One of the limitations of this study is that the analyses exclude individuals who have chosen not to hold a faculty position. In other words, only individuals who have chosen to try to balance work and family (as evidenced by their presence in the sample) are included in the analyses. Descriptive analyses suggest that comparable proportions of women and men junior faculty are in their prime childbearing years (under age 40) (about 36%). This suggests that women are no more likely than men to "opt out" of faculty careers because they want marriage and parenting to be a part of their lives. Nonetheless, because of the lack of information about qualified individuals who decide not to pursue faculty careers, the findings from this research may underestimate the relationship between family responsibilities and employment status.

A second limitation of this study pertains to the adequacy of the available variables in the database. Although the NSOPF:93 has many strengths, the database includes few direct measures of family responsibilities. Examples of important, but unavailable, variables include the ages of dependent children, timing of childbearing, childcare arrangements, employment status and occupation (e.g., academic or nonacademic) of the spouse, income of the spouse, amount of time spent out of the labor force because of family responsibilities, time devoted to household chores, and time devoted to child care. The NSOPF:93 also lacks measures of parent care-giving, a potentially important influence on employment status given that a 1997 survey found that 25% of the wage and salaried labor force nationwide had elder care responsibilities during the prior year, that employees with elder care responsibilities provided an average of 11 hours per week in assistance, and that 37% of employees with elder care responsibilities took time off from work to provide that assistance (Bond, Galinsky, & Swanberg, 1997). In addition, although holding the first or only job since earning the highest degree may be an inadequate proxy for mobility, the NSOPF:93 lacks appropriate alternative measures.

A third limitation pertains to the difficulty of determining the direction of causality between family responsibilities and employment status using this cross-sectional database. Therefore, this study focuses on exploring the relationship between family responsibilities and employment status among junior faculty, rather than on drawing conclusions about causality. To more fully explore the intercorrelations among the variables in the model, the analytic strategy involves entering conceptually related variables together. Sex and race are entered into the model first. Then family responsibilities are added to determine whether family responsibilities are related to employment status apart from the influence of human capital and structural characteristics. Measures of human capital are then entered into the model, followed by measures of structural characteristics.

Findings

Observed Relationships Between Family Responsibilities and Employment Status

The descriptive analyses suggest that women junior faculty hold a relatively higher proportion of full-time, nontenure-track positions and a relatively smaller proportion of full-time, tenure-track positions. Table 1 shows that women represent 52% of full-time, nontenure-track faculty, 47% of part-time, tenure-track faculty, 46% of part-time, nontenure-track faculty, and 43% of full-time, tenure-track faculty.

Regardless of employment status, a smaller proportion of women than men are married (67% versus 78% overall), but a higher proportion of women than men were previously married (17% versus 9% overall). Only 53% of women junior faculty have at least one child, compared with 70% of men junior faculty. A higher percentage of part-time, tenure-track faculty than of junior faculty overall are observed to have at least one child among both women (71% versus 53%) and men (83% versus 70%) junior faculty.

Relationship Between Family Responsibilities and Employment Status Controlling for Other Variables

Table 2 shows the odds-ratios for full-time, nontenure-track employment and part-time, nontenure-track employment relative to full-time, tenure-track employment among junior faculty. Columns 1 and 2 represent the "baseline" model that includes only measures of sex and race. Columns 3 and 4 show the relationship between family responsibilities and employment status controlling for differences in sex and race. Columns 5 and 6 show the odds-ratios when human capital characteristics are added to the model, and columns 7 and 8 show the odds-ratios when structural characteristics are also taken into account. The likelihood ratio test indicating the probability that all of the variables in the model are jointly equal to zero is rejected at the 0.1% level for all specifications.

The multinomial logistic regression analyses reveal that, controlling only for race, women are more likely to hold both full-time and part-time, nontenure-track positions than full-time, tenure-track positions (columns 1 & 2). Adding controls for family responsibilities does not change the relationship between sex and employment status (columns 3 & 4). Adding measures of human capital to the model eliminates the statistically significant relationship between sex and the likelihood of holding a part-time, nontenure-track position rather than a full-time, tenure-track position (column 6). Even after controlling for differences in race, family responsibilities, human capital, and structural characteristics, however, the odds of holding a full-time, nontenure-track position rather than a full-time, tenure-track position are higher for women than for men. These findings suggest that differences in race, family responsibilities, human capital, and structural characteristics do not fully account for the observed higher representation of women in full-time, nontenure-track positions.

Married faculty and previously married faculty are more likely than never married faculty to hold part-time, nontenure-track positions than full-time, tenure-track positions after controlling only for sex and race (column 4). Controlling also for human capital investment, however, eliminates these relationships (column 6). Contrary to expectations

TABLE 2

Odds-Ratios for Employment Status of Junior Faculty

Independent Variable	Sex and Race 1 Full-time, Nontenure Track	 2 Part-time, Nontenure Track	Family Responsibilities 3 Full-time, Nontenure Track	 4 Part-time, Nontenure Track	Human Capital 5 Full-time, Nontenure Track	 6 Part-time, Nontenure Track	Structural Characteristics 7 Full-time, Nontenure Track	 8 Part-time, Nontenure Track
Female	1.45***	1.15*	1.43***	1.15*	1.31**	1.03	1.32**	1.02
Black	0.97	0.65***	0.98	0.66***	0.91	0.63***	0.90	0.61***
Hispanic	0.69	0.82	0.70	0.83	0.62*	0.74	0.66	0.59**
Asian	0.74	0.54***	0.75	0.54***	1.07	1.06	0.97	1.08
Married			1.01	1.31**	0.95	1.11	1.00	1.08
Previously married			0.93	1.36**	0.87	1.04	0.96	0.95
At least one child			0.89	0.95	0.81*	0.76***	0.79*	0.83*
Less than doctorate					3.38***	6.35***	4.29***	6.29***
PhD from non-Research I					1.05	1.61***	1.11	1.57***
Research assistantship					0.78*	0.70***	0.75**	0.73**
Teaching assistantship					1.08	0.72***	1.07	0.67***
Years since highest degree					1.03***	1.05***	1.02***	1.05***
Years in current position					1.06***	1.05***	1.05***	1.06***
Only job since highest degree					1.24*	0.27***	1.12	0.27***
Other doctoral							0.69*	0.47***
Comprehensive I							0.65**	0.66***
Other 4-year							0.57***	0.48***
Specialized							0.24***	1.02
Two-year							0.69	0.71*
Unionized institution							1.15	1.07
Nursing & other health							1.16	0.65*
English & foreign languages							1.75**	2.11***
Education							1.24	1.57**

TABLE 2 (*Continued*)

Independent Variable	Sex and Race 1 Full-time, Nontenure Track	 2 Part-time, Nontenure Track	Family Responsibilities 3 Full-time, Nontenure Track	 4 Part-time, Nontenure Track	Human Capital 5 Full-time, Nontenure Track	 6 Part-time, Nontenure Track	Structural Characteristics 7 Full-time, Nontenure Track	 8 Part-time, Nontenure Track
Fine arts							0.68	1.46*
Psychology, sociology, other							1.36	2.57***
Other field							1.08	1.16
Biology							0.90	1.14
Mathematics & computer sci							1.74*	1.30
Business							1.27	1.06
Econ., pol. sci., history, law, philosophy							1.14	1.38
First-professional health							1.76*	0.98
Number of cases in analyses	6,270		6,270		6,270		6,270	
X^2, *df* (change −2 log likelihood)	66,	8***	83,	14***	1,900,	28***	2,360,	62***
Pseudo R^2 (Cox & Snell)	0.01		0.01		0.26		0.31	
Percent classified correctly	0.49		0.49		0.65		0.66	

Source: Analyses of 1993 National Study of Postsecondary Faculty (NSOPF:93).

Notes: Employment status is relative to full-time, tenure-track employment. Institutional type is relative to Research I university. Academic field is relative to engineering and physical science.

***$p < 0.001$. **$p < 0.01$. *$p < 0.05$.

based on human capital theory, junior faculty with at least one child are marginally ($p < 0.05$) less likely to hold both full-time and part-time, nontenure-track positions than full-time, tenure-track positions even after controlling for differences in other variables.

The analyses also reveal that educational attainment is an important predictor of employment status. The odds of holding either a full-time or part-time, nontenure-track position rather than a full-time, tenure-track position are substantially higher for faculty who have not earned a doctorate even after taking other differences into account. Holding a research assistantship during graduate school reduces the odds of holding either a full-time or part-time, nontenure-track position, net of other variables, while holding a teaching assistantship reduces the odds of holding a part-time, nontenure-track position. Junior faculty at non-Research I four-year colleges and universities appear to be less likely than junior faculty at Research I universities to hold full-time or part-time, nontenure-track positions. The odds of holding a part-time, nontenure-track position appear to be higher for faculty working in fields with among the highest proportions of women, English and foreign languages, education, fine arts, and psychology, sociology, and other social sciences. In contrast, faculty in the category with the highest proportion of women, nursing and non-first professional health, appear to be less likely to hold part-time, nontenure-track positions. These findings suggest that the concentration of women in particular academic fields may be related to the segregation of women by employment status.

To more fully explore sex differences in the relationship between family responsibilities and employment status, the analyses are repeated for women and men separately. The results, summarized in Table 3, show that, consistent with human capital theory, the odds of holding a part-time, nontenure-track position appear to be higher for married women than for other women even after controlling for race, human capital investment, and structural characteristics. Marital status appears to be unrelated to employment status among men junior faculty. Whereas having at least one child is unrelated to employment status for women junior faculty after controlling for other variables, men who have at least one child appear to be less likely than their childless male counterparts to hold a full-time, nontenure-track position.

Satisfaction with Nontenure-Track Employment Among Junior Faculty

Anecdotal evidence suggests that some fraction of nontenure-track faculty are content with their employment status because they prefer to spend their time on other well-paying jobs, hobbies, or raising children

TABLE 3

Odds-Ratios for Employment Status of Women and Men Junior Faculty: Fall 1992

	Full-time, Nontenure Track		Part-time, Nontenure Track	
Independent Variable	Women	Men	Women	Men
Black	0.81	1.01	0.45***	0.78
Hispanic	0.48	0.79	0.84	0.45***
Asian	1.17	0.95	1.19	1.00
Married	1.09	0.95	1.41*	0.78
Previously married	0.85	1.11	1.10	0.84
At least one child	0.99	0.62**	0.90	0.82
Less than doctorate	4.12***	4.44***	8.34***	5.29***
PhD from non-Research I	1.42	0.91	2.09***	1.29
Research assistantship	0.64**	0.84	0.65**	0.82
Teaching assistantship	1.10	1.05	0.63***	0.68***
Number of years since highest degree	1.00	1.04***	1.04***	1.06***
Number of years in current position	1.08***	1.04**	1.05***	1.06***
Only job since highest degree	0.84	1.42*	0.31***	0.21***
Unionized institution	0.97	1.31*	0.88	1.29**
Other doctoral	0.55**	0.89	0.45***	0.53***
Comprehensive I	0.43***	0.98	0.61**	0.73*
Other 4-year	0.31***	1.08	0.40***	0.58**
Two-year	0.14***	0.45**	0.85	1.33
Specialized institution	0.50*	0.92	0.94	0.61*
Nursing & other health	1.30	0.93	0.68	1.03
English & foreign languages	2.04	1.79*	3.07**	1.62*
Education	1.26	1.52	2.29*	1.21
Fine arts	1.02	0.48*	2.39*	1.15
Psychology, sociology, other	1.57	1.34	3.60***	2.29***
Other field	1.03	1.14	1.18	1.30
Biology	1.13	0.80	1.69	1.00
Mathematics & computer science	1.91	1.76*	1.24	1.49
Business	1.52	0.83	1.29	1.21
Econ., pol. sci., history, law, philosophy	1.55	0.93	1.84	1.28
First-professional health	1.56	2.00*	0.89	1.12
Number of cases in the analyses	2,902	3,368		
R^2, *df* (change -2 log likelihood)	1,081, 60***	1,407, 60***		
Pseudo R^2 (Cox & Snell)	0.311	0.341		
Percent classified correctly	66%	68%		

SOURCE: Analyses of 1993 National Study of Postsecondary Faculty (NSOPF:93).

NOTES: Employment status is relative to full-time, tenure-track employment. Institutional type is relative to Research I university. Academic field is relative to engineering and physical science.

***$p < 0.001$. **$p < 0.01$. *$p < 0.05$.

(Flynn, Flynn, Grimm & Lockhart, 1986; Wilson, 1998). Exploring the extent to which junior faculty are satisfied holding part-time and nontenure-track positions is limited by the variables available in the NSOPF:93 database. For example, part-time faculty, but not full-time, nontenure-track faculty, were asked their reasons for their current employment status. Descriptive analyses of the available data reveal that women and men are equally likely to report holding part-time, nontenure-track positions because they prefer to work part-time (about 56%). Table 4 shows that women are more likely than men to report holding a part-time, nontenure-track position because a full-time position is unavailable (49% versus 38%) but less likely to report holding a part-time, nontenure-track position because they are supplementing their incomes (48% versus 61%). These data suggest that, despite a generally stated preference for working part-time, women may actually be less likely than men to hold part-time, nontenure-track positions because they have voluntarily chosen such status.

Another approach to understanding the extent to which junior faculty are satisfied holding nontenure-track positions is to examine the relative importance of various characteristics in a decision to leave the current job. The descriptive data presented in Table 5 suggest that, among both women and men, the prospect of a tenured position is somewhat less important for faculty holding nontenure-track positions than for faculty holding full-time, tenure-track positions (about 47% versus 64%). Opportunity for advancement and job security are very important for most faculty, although both appear to be somewhat less important for part-time, nontenure-track faculty than for full-time, tenure-track faculty. Despite these differences, the data suggest that a substantial portion of women and men nontenure-track faculty are interested in a tenured position, opportunity for advancement, and job security. The importance of some characteristics that are likely associated with family responsibilities, such as geographic location and schools for children, does not appear to vary by employment status. Among both women and men, the

TABLE 4

Reasons Part-time, Nontenure-Track Faculty Are Working Part-time: Fall 1992

Characteristic	Total	Women	Men	Statistical Difference
Total	100%	100%	100%	
Preferred part-time	56%	55%	57%	n.s.
Full-time unavailable	43%	49%	38%	***
Supplementing income	55%	48%	61%	***
To be in academic environment	71%	71%	71%	n.s.
Finishing graduate degree	9%	11%	7%	***
Other reason	20%	20%	20%	n.s.

Source: Analyses of 1993 National Study of Postsecondary Faculty (NSOPF:93).
Note: n.s. indicates not statistically significant.
***$p < 0.001$. **$p < 0.01$. *$p < 0.05$.

availability of a job for the spouse appears to be somewhat less important for part-time, nontenure-track faculty than for full-time faculty. Regardless of employment status, however, spousal employment appears to be a more important concern for women than for men. Nonetheless, both women and men junior faculty appear to be satisfied with their choice of profession regardless of employment status. Table 6 shows that, on average, both women and men junior faculty generally agree that they would choose an academic career again.

Conclusions and Implications

Although Bowen and Schuster (1986) predicted that differences in the status of, and rewards received by, women and men faculty would diminish as the number of women entering academic careers continued to increase, the findings from this study show that sex differences continue to exist in employment status. Even after controlling for differences in race, family responsibilities, human capital, and structural characteristics, women are more likely than men to hold full-time, nontenure positions, positions of lower status in the academic labor market hierarchy.

TABLE 5

Percent of Junior Faculty Reporting Various Characteristics to be Very Important in the Decision to Leave Current Job: Fall 1992

Characteristic	Sex	Total	Full-time, Tenure Track	Full-time, Nontenure Track	Part-time, Nontenure Track
Tenured position	Total	51%	64%	47%	42%
	Women	52%	65%	48%	44%
	Men	50%	64%	46%	40%
Opportunity advancement	Total	61%	66%	62%	57%
	Women	63%	69%	65%	59%
	Men	59%	64%	59%	55%
Job security	Total	68%	74%	72%	63%
	Women	72%	76%	77%	69%
	Men	64%	72%	66%	57%
Geographic location	Total	59%	61%	58%	58%
	Women	61%	63%	60%	61%
	Men	58%	60%	56%	55%
Schools for children	Total	54%	58%	53%	52%
	Women	53%	53%	54%	52%
	Men	55%	62%	52%	51%
Job for spouse	Total	44%	51%	48%	37%
	Women	49%	56%	57%	42%
	Men	39%	48%	39%	33%

SOURCE: Analyses of 1993 National Study of Postsecondary Faculty (NSOPF:93).

TABLE 6

Level of Agreement that a Junior Faculty Member Would Choose an Academic Career Again: Fall 1992

Sex	Total	Full-time, Tenure Track	Full-time, Nontenure Track	Part-time, Nontenure Track
Women	3.43	3.45	3.40	3.40
Men	3.45	3.46	3.42	3.45

SOURCE: Analyses of 1993 National Study of Postsecondary Faculty (NSOPF:93).
NOTE: Scale is from 1 to 4, with 1 indicating strongly disagree and 4 indicating strongly agree.

Though both human capital and structural approaches were shown to be useful for understanding the distribution of faculty by employment status, the results of this research suggest that human capital and structural approaches to the academic labor market do not fully account for the concentration of women in full-time, nontenure-track positions. One possible explanation for this finding is that the model omitted, or inadequately measured, important aspects of human capital and structural characteristics, as described in the limitations section. A second possible explanation is that women prefer to hold full-time, nontenure-track positions for reasons that are not adequately captured by the available proxies for family responsibilities. The finding that about 55% of women with part-time, nontenure-track appointments prefer to work part-time may be consistent with this explanation. Alternatively, women may be more likely to hold these lower status positions because they are perceived by colleges and universities to be less productive and/or incapable of succeeding in full-time, tenure-track positions. The descriptive data showing that a higher share of women than men are working part-time because a full-time position is unavailable may support the appropriateness of this explanation.

The findings from this study also suggest that the employment of women in nontenure-track positions is attributable in part to their marital and parental status. Although a smaller share of women than men junior faculty are married (67% versus 78%), being married increases the odds of holding a part-time, nontenure-track position for women but not for men. While a smaller share of women than men junior faculty have at least one child (53% versus 70%), having at least one child reduces the odds of holding a full-time, nontenure-track position for men but is unrelated to employment status for women. As Toutkoushian (1998) noted, these sex differences in employment status for women and men faculty may be attributable to either differences in the supply of women and men faculty who are married or parents or to differences in the demand for faculty who are married or parents. Regardless, while some research (e.g., Bellas, 1992; Toutkoushian, 1998) has shown that married men faculty benefit from having wives in terms of their productivity and salaries, this study suggests that married men faculty and men faculty with children are also benefiting from their marital and parental status in terms of their employment status.

Because the NSOPF:93 lacks variables describing the nature of the spouse's employment, future research should examine the extent to which married women are more likely to hold part-time, nontenure-track positions because they are married to other academics. Using a national survey of college and university faculty in 1989, Astin and Milem (1997) showed that a higher share of married women faculty than married men faculty are married to other academics (40% versus 35%). Whereas some research suggests that women with academic spouses may be benefitting in terms of their productivity, rank, and salaries possibly because of greater access to collegial networks (Astin & Milem, 1997), this study suggests that married women—a substantial proportion of whom are likely to have academic spouses—may be disadvantaged with regard to their employment status because of a lack of mobility.

On the surface, the descriptive analyses suggest that a substantial portion of women and men junior faculty are relatively satisfied holding lower status (i.e., nontenure-track) positions. Regardless of employment status, both women and men generally agreed that they would choose an academic career again. Nonetheless, future research should also further explore the satisfaction of women and men nontenure-track faculty, particularly given the conclusion of Tack and Patitu (1992) that many married women faculty may be "diluting their professional ambitions and assuming part-time faculty positions" because the demands of work and family are too overwhelming (p. 53).

To some extent, the participation of women in the labor force may always be limited by family responsibilities (Hough, 1987). Even though the share of men who are assuming care-giving responsibilities is growing (Bond et al., 1997), the effects of marital and parental responsibilities on faculty careers are likely to continue to be greater for women than men (Gappa & Leslie, 1993). Dual career relationships typically require one or both partners to make sacrifices regarding career opportunities, particularly in terms of where to live (Gappa & Leslie, 1993). Marriage is likely to impose a greater hardship on the career development and advancement of women than men because a higher percentage of employed women than of employed men are in dual career marriages (89% versus 69%) (Bond et al., 1997). Moreover, even in the 1990s, many families may focus on maximizing the husband's rather than the wife's employment status (Marwell et al., 1979). McElrath (1992) found that, among criminology faculty, women were three times as likely to interrupt their careers because of reasons related to their husband's employment than for maternity. After controlling for education, experience, publications, and parental status, a career disruption and the number of job changes were associated with a lower probability of tenure for women and, among tenured women, a longer time to tenure. In contrast, career disruptions and job changes were unrelated to either the probability of being tenured or the number of years to tenure among men (McElrath, 1992). The findings from this study further suggest that the effects of family responsibilities are less advantageous for women than for men.

Some critics of human capital theory have argued that "many workers who could contribute to the economic advance of the nation have been confined to low-status jobs where they are not allowed to be productive" (DeYoung, 1989, p. 161). Because of the challenges Youn (1992) identifies with moving across segments of the academic labor market, faculty who hold nontenure-track appointments but who aspire to tenure-track or tenured appointments are unlikely to achieve their goal. A number of researchers (e.g., Bowen & Schuster, 1986; Chronister, Baldwin & Bailey, 1996; Franklin, Laurence, & Denham, 1988; Gappa & Leslie, 1993; Kasper, Bronner, Gray, Kreiser, & Rosenthal, 1986; Lomperis, 1990; Rajagopal & Farr, 1992;) have concluded that part-time and nontenure-track faculty generally receive less encouragement and support for research activities, as evidenced by their less desirable teaching assignments and heavier teaching loads, lack of collegial support, and lack of access to resources for research including release time, funding, and facilities.

From an institutional perspective, the use of nontenure-track faculty may appear to have some financial and programmatic advantages. Nontenure-track appointments may provide colleges and universities with greater flexibility to respond to enrollment changes and shifts in enrollment across academic disciplines and may enable them to offer specialized courses without the commitment of resources that a tenure-track appointment entails. Nonetheless, Kasper et al. (1986), Franklin et al. (1988), and others have argued that the increasing use of part-time and nontenure-track faculty undermines academic standards and diminishes the quality of undergraduate education. Gappa and Leslie (1993) concluded, based on their examination of the use of part-time faculty at 18 colleges and universities, that using part-time faculty does not necessarily improve efficiency or cost effectiveness. As an example, the use of part-time and nontenure-track faculty typically requires that tenured and tenure track full-time faculty assume the burden of student advising, committee work, and other activities in which nontenure-track faculty do not fully participate.

Although some (e.g., Franklin et al., 1988) have recommended that some nontenure-track positions be converted to tenure-track assistant professor appointments, nontenure-track faculty are likely to continue to comprise a substantial proportion of our nation's faculty (Gappa & Leslie, 1993). The findings from this research suggest that individual colleges and universities should reexamine their policies and procedures regarding recruitment and tenure. With regard to recruitment, institutions should ensure that the tendency of women to hold full-time, nontenure-track positions rather than full-time, tenure-track positions is attributable to a genuine preference for such positions. Colleges and universities should also examine their policies and procedures regarding tenure to ensure that women are not pressured to choose between a tenure-track position and motherhood. This study showed that only 2% of junior faculty nationwide hold part-time, tenure-track positions. Of 191 colleges and universities in one survey, only 11% had a policy offering tenure to part-time faculty (Raabe, 1997). By creating flexibility in the tenure process (e.g., by allowing part-time faculty to pursue tenure), and by ensuring that all administrators and faculty understand the need for such policies (Finkel et al., 1994), colleges and universities will help to establish a more "family friendly" environment.

As others (e.g., Marshall & Jones, 1990; Sorcinelli & Near, 1989) have suggested, all faculty would benefit from institutional efforts that support faculty in managing work and family roles. The Families and Work Institute concluded from its 1997 survey of employees in a variety of occupations nationwide that a supportive workplace environment is critical to the effectiveness, satisfaction, commitment, and retention of workers regardless of industry (Bond et al., 1997). A study of faculty at one university suggests that job and life satisfaction are more highly correlated among college and university faculty than among the general population (Sorcinelli & Near, 1989). Both married men and women with children are concerned about dual careers, commuter marriages, and childrearing (Sorcinelli & Near, 1989). Therefore, all faculty would likely benefit from such initiatives as workshops on time and stress management and sex role socialization, supportive counseling (especially during family and career changes), higher quality and more available childcare, and employment assistance for spouses and partners, as well as more flexible leaves and sabbaticals. Nonetheless, a 1991 survey of 191 colleges and universities showed that, while most institutions had a policy regarding unpaid or paid leave for mothers at childbirth, less than one-half had policies covering job assistance for the spouse, accommodative scheduling, unpaid leave for fathers at childbirth, or on-campus childcare centers (Raabe, 1997). Moreover, even when such policies as accommodative scheduling and job sharing are in place, they are reported to be only rarely used (Raabe, 1997). By adopting and encouraging the use of policies, practices, and initiatives that recognize that many faculty are also spouses and parents, colleges and universities will create an environment that fosters the success of both women and men faculty.

References

Astin, H. S., & Bayer, A. E. (1979). Pervasive sex differences in the academic reward system. In D. R. Lewis & W. E. Becker (Eds.), *Academic Rewards in Higher Education* (Chap. 10). Cambridge, MA: Balinger.

Astin, H. S., & Davis, D. E. (1985). Research productivity across the life and career cycles: Facilitators and barriers for women. In M. F. Fox (Ed.), *Scholarly writing and publishing: Issues, problems, and solutions.* Boulder, CO: Westview Press.

Astin, H. S., & Milem, J. F. (1997). The status of academic couples in U.S. institutions. In M. A. Ferber & J. W. Loeb (Eds.), *Academic couples: Problems and Promises* (pp. 128–155). Urbana: University of Illinois Press.

Austin, A. E., & Pilat, M. (1990). Tension, stress, and the tapestry of faculty lives. *Academe, 77,* 38–42.

Barbezat, D. (1988). Gender differences in the academic reward system. In D. W. Breneman & T. I. Youn (Eds.), *Academic labor markets and careers* (pp. 138–164). New York: The Falmer Press.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy, 70 Supplement* (5), 9–49.

Becker, G. S. (1985). Human capital, effort, and the sexual division of labor. *Journal of Labor Economics, 3*(1), S33–S58.

Becker, G. S. (1993). *Human capital.* Chicago: University of Chicago Press.

Bellas, M. L. (1992). The effects of marital status and wives' employment on the salaries of faculty men: The (house) wife bonus. *Gender & Society, 6,* 609–622.

Bellas, M. L. (1994). Comparable worth in academia: The effects on faculty salaries of the sex composition and labor-market conditions of academic disciplines. *American Sociological Review, 59,* 807–821.

Bellas, M. L. (1997a). Disciplinary differences in faculty salaries: Does gender bias play a role? *Journal of Higher Education, 68,* 299–321.

Bellas, M.L. (1997b). The scholarly productivity of academic couples. In M. A. Ferber & J. W. Loeb (Eds.), *Academic couples: Problems and promises.* Urbana: University of Illinois Press.

Bellas, M. L., & Toutkoushian, R. K. (1999). Faculty time allocations and research productivity: Gender, race, and family effects. *Review of Higher Education, 22,* 367–390.

Bielby, D. D., & Bielby, W. T. (1988). She works hard for the money: Household responsibilities and the allocation of work effort. *American Journal of Sociology, 93,* 1031–1059.

Bond, J. T., Galinsky, E., & Swanberg, J. E. (1997). *The 1997 national study of the changing workforce.* New York: Families and Work Institute.

Bowen, H. R., & Schuster, J. H. (1986). *American professors: A national resource imperiled.* New York: Oxford University Press.

Bronstein, P., Rothblum, E. D., & Solomon, S. E. (1993). Ivy halls and glass walls: Barriers to academic careers for women and ethnic minorities. In J. Gainen & R. Boice (Eds.), *Building a diverse faculty* (Vol. 53, pp. 17–31). San Francisco: Jossey-Bass Publishers.

Chronister, J. L., Baldwin, R. G., & Bailey, T. (1996). Full-time non-tenure track faculty: Current status,.condition, and attitudes. In D. E. Finnegan, D. Webster, & Z. F. Gamson (Eds.), *Faculty and faculty issues in colleges and universities* (2nd ed., pp. 326–336). Needham Heights, MA: Simon & Schuster Custom Publishing, ASHE Reader Series.

Chronister, J. L., Gansneder, B. M., Harper, E., & Baldwin, R. G. (1997). Full-time non-tenure track faculty. *NEA Higher Education Research Center Update, 3*(5), 1–4.

Cole, J. R., & Zuckerman, H. (1987). Marriage, motherhood, and research performance in science. *Scientific American, 256*(2), 119–125.

Cooney, T. M., & Uhlenberg, P. (1989). Family-building patterns of professional women: A comparison of lawyers, physicians, and postsecondary teachers. *Journal of Marriage and the Family, 51,* 749–758.

Creamer, E. G. (1998). *Assessing faculty publication productivity: Issues of equity.* (ASHE-ERIC Higher Education Report Vol. 26, No. 2). Washington, DC: The George Washington University: Graduate School of Education and Human Development.

Dey, E. L. (1994). Dimensions of faculty stress: A recent survey. *Review of Higher Education, 17,* 305–322.

DeYoung, A. J. (1989). *Economics and American education: A historical and critical overview of the impact of economic theories on schooling in the United States.* White Plains, NY: Longman Inc.

Dreijmanis, J. (1991). Higher education and employment: Is professional employment a right? *Higher Education Review, 23*(3), 7–18.

England, P. (1982). The failure of human capital theory to explain occupational sex segregation. *Journal of Human Resources, 17,* 358–370.

Ferber, M. A., & Hoffman, E. P. (1997). Are academic partners at a disadvantage? In M. A. Ferber & J. W. Loeb (Eds.), *Academic couples: Problems and promises.* Urbana: University of Illinois Press.

Finkel, S. K., Olswang, S., & She, N. (1994). Childbirth, tenure, and promotion for women faculty. *Review of Higher Education, 17,* 259–270.

Flynn, E. A., Flynn, J. F., Grimm, N., & Lockhart, T. (1986). The part-time problem: Four voices. *Academe, 72*(1), 12–18.

Franklin, P., Laurence, D., & Denham, R. D. (1988). When solutions become problems: Taking a stand on part-time employment. *Academe, 74*(3), 15–19.

Gappa, J., & Leslie, D. (1993). *The invisible faculty: Improving the status of part-timers in higher education.* San Francisco: Jossey-Bass.

Hamovitch, W., & Morgenstern, R. D. (1977). Children and the productivity of academic women. *Journal of Higher Education, 48,* 633–645.

Hough, J. R. (1987). *Education and the national economy.* New York: Croom Helm.

Johnson, G. E., & Stafford, F. P. (1974). The earnings and promotion of women faculty. *American Economic Review, 64,* 888–903.

Kasper, H., Bronner, F., Gray, M. W., Kreiser, B. R., & Rosenthal, J. R. (1986). 1986 report on full-time non-tenure track appointments. *Academe, 72*(4), 14a-19a.

Kasten, K. L. (1984). Tenure and merit pay as rewards for research, teaching, and service at a research university. *Journal of Higher Education, 55,* 500–514.

Kirshstein, R. J., Matheson, N., & Jing, Z. (1997). *Instructional faculty and staff in higher education institutions: Fall 1987 and Fall 1992.* Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES 97–447).

Korenman, S., & Neumark, D. (1991). Marriage, motherhood, and wages. *Journal of Human Resources, 27,* 233–255.

Lomperis, A. M. T. (1990). Are women changing the nature of the academic profession? *Journal of Higher Education, 61,* 643–677.

Marshall, M. R., & Jones, C. H. (1990). Childbearing sequence and the career development of women administrators in higher education. *Journal of College Student Development, 31,* 531–537.

Marwell, G., Rosenfeld, R., & Spilerman, S. (1979). Geographic constraints on women's career in academia. *Science, 205,* 1225–1231.

McElrath, K. (1992). Gender, career disruption, and academic rewards. *Journal of Higher Education, 63,* 269–281.

Menard, S. (1995). *Applied logistic regression analysis.* Thousand Oaks: Sage.

Polachek, S. W. (1977). Occupational segregation among women: Theory, evidence, and a prognosis. In C. B. Lloyd, E. S. Andrews, & C. L. Gilroy (Eds.), *Women in the Labor Market.* New York: Columbia University Press.

Raabe, P. H. (1997). Work-family policies for faculty: How "career- and family-friendly" is academe? In M. A. Ferber & J. W. Loeb (Eds.), *Academic couples: Problems and Promises* (pp. 208–225). Urbana: University of Illinois Press.

Rajagopal, I., & Farr, W. D. (1992). Hidden academics: The part-time faculty in Canada. *Higher Education, 24,* 317–331.

Riemenschnieder, A., & Harper, K. V. (1990). Women in academia: Guilty or not guilty? Conflict between caregiving and employment. *Initiatives, 53*(2), 27–35.

Rosenfeld, R. A., & Jones, J. A. (1987). Patterns and effects of geographic mobility for academic women and men. *Journal of Higher Education, 58,* 493–515.

Smart, J. C. (1991). Gender equity in academic rank and salary. *Review of Higher Education, 14,* 511–526.

Smart, J. C., & McLaughlin, G. W. (1978). Reward structures of academic disciplines. *Research in Higher Education, 8,* 39–55.

Sorcinelli, M. D., & Near, J. P. (1989). Relations between work and life away from work among university faculty. *Journal of Higher Education, 60,* 59–81.

Sorenson, E. (1989). Measuring the effect of occupational sex and race composition on earnings. In R. T. Michael, H. I. Hartmann, & B. O'Farrell (Eds.), *Pay equity: Empirical inquiries* (chap. 2). Washington, DC: National Academy Press.

Tack, M. W., & Patitu, C. L. (1992). *Faculty job satisfaction: Women and minorities in peril.* (ASHE-ERIC Higher Education Report No. 4). Washington, DC: The George Washington University, School of Education and Human Development.

Toutkoushian, R. K. (1998). Racial and marital status differences in faculty pay. *Journal of Higher Education, 69,* 513–541.

Wilson, R. (1998, July). For some adjunct faculty members, the tenure track holds little appeal. *Chronicle of Higher Education,* pp. A9–A10.

Youn, T. I. K. (1988). Studies of academic markets and careers: An historical review. In D. W. Breneman, & T. I. K. Youn (Eds.), *Academic labor markets and careers* (chap. 1, pp. 8–27). New York: The Falmer Press.

Youn, T. I. K. (1992). The sociology of academic careers and academic labor markets. *Research in Labor Markets, 13,* 101–130.

Review Essay

The Role of Faculty in Assessing Colleagues' Teaching
The Peer Review of Teaching: A Sourcebook,
by Nancy Van Note Chism.
Bolton, MA: Anker, 1999. 141 pp., $24.95 (paper)

LARRY KEIG, The University of Northern Iowa

For well over twenty years, peer review of college teaching has been promoted in the exhortative literature, justified in theory, and supported by reports of experimental and qualitative research. Yet, except for the faculty's role in personnel decision making, peer review is rarely practiced; and even when and where it has been put into place (interestingly, mostly at the departmental level of research universities, where results sometimes have been reported in journals), it hardly ever has sustained itself for any appreciable length of time. In *Peer Review of Teaching,* Chism suggests that one reason for its infrequent practice is a lack of guiding principles and materials for developing and implementing it.

Chism compellingly argues for the involvement of faculty in assessing colleagues' teaching and—probably most significantly—provides a wide range of useful materials for peer review of many teaching-related activities. She clearly describes the potential value of peer review for both improving teaching (formative evaluation) and making decisions regarding reappointment, tenure, promotion, and merit compensation (summative evaluation). She also persuasively argues not only for employing it alongside student ratings of courses and instructors but also for other forms of assessment, recognizing that credible efforts to improve and assess teaching require a comprehensiveness not often found

Larry Keig is assistant professor, Department of Educational Leadership, Counseling, and Postsecondary Education.

The Journal of Higher Education, Vol. 72, No. 5 (September/October 2001)
Copyright © 2001 by The Ohio State University

in college faculty evaluation. She unquestionably provides the best and most extensive collection of tools for the peer review of teaching to date.

The book is in two parts. The four chapters comprising Part 1 provide an overview: Developing a Rationale and Understanding of Peer Review (Chapter 1), Setting Up a System of Peer Review (Chapter 2), Understanding Roles and Goals of a Peer Review System (Chapter 3), and Identifying the Focus of a Peer Review of Teaching System (Chapter 4). Five chapters, which correspond to different methods of peer review, make up Part 2: Peer Review of Course Materials (Chapter 5), Classroom Observation (Chapter 6), Leadership for Teaching: Contributing to Scholarship of Teaching and Departmental Teaching Efforts (Chapter 7), Teaching Portfolios (Chapter 8), and Summary (Chapter 9). Chapters 5 through 9 include the practical "how to" information (resources and forms) that practitioners could use in assessing colleagues' teaching performance; this is what readers who want to know what and how to peer review should find especially useful.

Chism almost certainly would support Centra's contention that "unless faculty members are willing to leave the evaluation of teaching to students, who possess only a limited view, or to administrators, who don't have the time or necessary background, then they must be willing to invest their time in efforts at peer evaluation of teaching" (1986, p. 1). Extending the argument for peer review of teaching further, Chism maintains that it is "the responsibility of the profession to monitor itself, . . . to make teaching 'community property'; . . . that the evaluation of teaching must reflect the complexity of teaching itself; . . . [and] that in order for teaching to be valued, there must be an accepted way to judge it" (p. 6). For the professoriate to take teaching (and, by extension, student learning) seriously, for assessment to reflect teaching's complex nature, and for teaching to be valued as much as research and scholarship (traditionally defined), she insists it must be assessed by multiple sources (peers as well as students and administrators), by a variety of methods (at least evaluation of course materials, leadership for teaching, and teaching portfolios as well as classroom observation), and for at least two different, though complementary purposes (improving teaching as well as making of personnel decisions). She also fairly addresses objections to peer review (e.g., the sanctity of the classroom, the distinctions between peer and colleague, vulnerability of peer reviewers, the time factor, lack of standards of effective teaching and the possible disjunction between "effective teaching" and student learning, validity and reliability issues, and undesirable aftereffects).

In emphasizing resources (i.e., forms for evaluating teaching), Chism may unwittingly promote products more appropriate for summative evaluation than for formative assessment. In saying that the resources in the sourcebook "might be a helpful starting point for the discussion of standards [of effective teaching and that] they should be adapted and modified as the faculty sees fit," (p. 21), however noble the intent, some of the resources she provides may run counter to an approach often recommended for communicating feedback to faculty in improving teaching. Centra has suggested, for example, that for some types of formative peer review qualitative methods may be more valuable than quantitative findings, noting that "a qualitative approach would involve descriptions of classroom instruction based on the perceptions of the observers. Rating scales and numerical judgments would not be necessary or useful. In many instances, in fact, the scales are not appropriate for all styles of teaching. But descriptions by several observers will more likely reflect possible biases and the resulting narrative could be much more useful" (Centra, 1986, pp. 2–3). Even though Chism has suggested, in directions for use on most of the forms, that peer reviewers focus on producing comments, reviewers may be tempted, given time constraints, to rate rather than narrate, to judge rather than describe. Unlike summative review, formative evaluation should include nonjudgmental descriptions of faculty members' teaching, not only by colleagues but also by academic administrators (especially, as Chism perceptively notes, by the department chair) and, where available, teaching consultants.

Chism offers a particularly useful framework for identifying roles faculty should play in assessing peers' teaching: first-order and second-order, noting that both are essential. First-order roles are assessments of "those things that peers are most able to judge [for themselves]." Second-order roles are their reviews of "judgments and evidence submitted by others" (p. 35). In identifying roles in each of these categories, Chism probably relies too heavily on a classification provided by Cohen and McKeachie (1980), good as it is, and not extensively enough on other useful lists and discussions (e.g., Batista, 1976; Scriven, 1980, 1985, 1987; and Soderberg,1986). The roles recommended by all of these scholars (and others) are summarized by Keig and Waggoner (1994, pp. 30–34; 1995, pp. 55–57).

Soderberg's (1986) three-dimensional framework for evaluating college teaching is also useful for thinking about how peer review might contribute to better teaching and more meaningful student learning. According to Soderberg, faculty members' decision making and performance can be assessed by knowledgeable constituencies on their pre-interactive decision making, on how effectively they interact with students, and on how well they "capture the process nature of instruction," including the feedback they and their students receive and act upon (Soderberg, 1986, p. 16). Chism has designed materials for peer reviewers at all three of these stages, focusing primarily on pre-interactive decision making and interaction rather than on, and somewhat at the expense of, process indicators.

Her resources for peer review of course materials, probably best used as guides for narratives on performance and/or conversations between or among faculty, are especially comprehensive and well conceived; in much of this section of the book, she has broken new ground. These resources include assessments of the following: syllabi; ground rules for discussions and other policy and/or procedure documents; course packet and textbook content; course bibliographies; overhead transparencies or presentation slides; course handouts; multimedia course materials; tests written by faculty and administered to students; and class assignments and exercise sheets. They may be less suitable for summative evaluation, where a more global assessment, focusing on a limited number of items, probably is more relevant to those making personnel decisions (McKeachie, 1994, pp. 329–330).

Chism's analysis of the strengths and limitations of classroom observation is probing, and the resources she provides for peer observation capture important elements of interaction between teacher and student. Because peer observation may not be reliable enough to use for summative evaluation (Centra, 1993, 1986; McKeachie, 1994), the resources probably are most suitable for formative peer review. If used for improving teaching, the forms might be best employed as guides for written or verbal non-evaluative feedback of how teachers interact with students. Peer reviewers may find the Narrative Log (pp. 82–84) especially useful because it can "help instructors review a class after it has occurred. They can stimulate recall and freeze the class in time for the purpose of examination . . . [of] the fit of actions to goals, student learning issues, alternative ways situations could have been handled, and the like" (p. 82).

As Cross has noted, "The ultimate criterion of effective teaching is effective learning. There is simply no other reason for teaching. . . . Good teaching is not so much of a performing act as an evocative process. The purpose is to involve students actively in their own learning and to elicit from them their best learning performance (1991, p. 20). In order for students to perform at optimum levels, they need, among other things, to receive meaningful feedback from faculty on their performance. Acknowledging the significance of instructor feedback, Chism has provided a valuable resource for assessing instructor comments on student work, but it regrettably is tucked among the resources for assessing course materials. Because course materials and instructor feedback come at different ends of the teaching process, this resource probably should be one of several in a chapter devoted to peer assessment of faculty evaluation of students' academic work. Although the form Chism provides for assessing instructor comments on student work is generic enough for assessing various types of student work, it seems most appropriate for comments on written assignments. Resources also could be provided for assessing instructor comments on classroom participation, oral communication assignments, cooperative and group learning projects, technology-based projects, student portfolios, inquiry-based projects, experiential learning, and the like. Information provided by chapter authors in Anderson and Speck's 1998 edited book, *Changing the Way We Grade Student Performance: Classroom Assessment and the New Learning Paradigm*, could be used as a springboard for developing additional resources for assessing instructor comments of student work. Chism should not be faulted, though, for not including more resources of this type, because this is largely uncharted territory; however, resources of this type are desperately needed if higher education is to make the shift from the instruction paradigm to the learning paradigm that Barr and Tagg (1995) (and others, in one way or another) have advocated.

In recent years, it has become fashionable to invoke references to *scholarship of teaching* and *teaching portfolios* in discussions of teaching, learning, and assessment. Ernest Boyer's *Scholarship Reconsidered: Priorities of the Professoriate* (1990) has become a tremendously influential, if not fully understood, source in discussions on these issues. College faculty members who consider themselves *teachers* have embraced the premise that professors should be (or become) scholars of teaching because faculty cannot possibly learn everything they need to know about how to teach effectively, and how students learn best and are motivated, from their experiences in classrooms. Yet largely unaddressed (if addressed at all) in discussions on the scholarship of teaching are the means by which professors come to understand the knowledge bases of teaching, learning, and assessment. As in every other scholarly field, the body of literature in each of these areas is extensive but remains largely unknown to most faculty. Using Shulman's broad categories of the knowledge base of teaching as but one example, a teaching scholar would demonstrate *depth* of understanding not only of content in the teaching field but also in general pedagogical knowledge, pedagogical knowledge specific to the content area, curricular knowledge, knowledge of learners, knowledge of educational contexts, and "knowledge of educational ends, purposes, and values, and their philosophical and historical grounds" (Shulman, 1987, p. 8).

Chism's description of the scholarship of teaching—"the activities through which faculty explore conceptions central to the teaching of their field, assess the effects of different teaching strategies on student learning in the discipline, and pose new directions for exploration" (p. 100)—captures the essence of the scholarship of teaching. But her discussion of and resources for the peer review of it appear somewhat superficial, mainly because they are not tied closely enough to the knowledge bases of effective teaching, student learning and motivation, and assessment. (Criticism of this type could also be leveled against most of what has been written on assessing the scholarship of teaching, for the same reasons.) In this same chapter, Chism includes resources for teaching leadership efforts and for nonclassroom teaching activities. These resources are presented in the form of thoughtful questions which could lead to meaningful dialogue among faculty on significant teaching and learning issues and for a peer reviewer's written analyses of another's performance.

In a teaching portfolio, faculty have opportunities to document their teaching philosophy, what and how they teach, the instructional materials they design and use, their students' work along with their comments on and evaluation of it, and how their efforts have been received by students, colleagues, and administrators. In addition, as Chism astutely notes, "A good portfolio is woven together by narrative commentary from the faculty member that describes the context for the documentation and presents reflections on the teaching self. It presents multiple sources of evidence, chronicles the development of the instructor, and projects a future vision" (p. 108). She provides sound information not only on what should be included in a portfolio but also on how it might be peer reviewed. Peer reviewers should find both the insightful questions she lays out for formative assessment and the solid, more global items she includes for summative evaluation useful.

Teaching portfolios have been a topic of discussion for improving and evaluating teaching since the 1980s. Discussion on how *course* portfolios might be used for the same purposes has begun more recently. Even though Chism states that "teaching portfolios can concentrate on only one course [as well as] span much longer time periods" (p. 108), it is somewhat surprising she has not integrated the literature on course portfolios with that on teaching portfolios, especially because, as Shulman has pointed out, "The course portfolio is a central element in the argument that teaching and scholarship are neither antithetical nor incompatible, . . . [and that a course], in its design, enactment, and analysis, is as much an act of inquiry and invention as any other activity more traditionally called 'research.'" (1999, p. 5). Chism's argument for a scholarship of teaching probably would be strengthened if she probed what is known about course, as well as teaching, portfolios.

The Peer Review of Teaching is not only a valuable addition to the literature on assessment and evaluation but also a significant resource for practitioners looking for more powerful means for reviewing teaching and for improving the quality of student learning. It clearly goes beyond a rehash of exhortative and theoretical literature and research studies. In this book, Chism provides the best materials currently available for assessing course materials and valuable resources for peer classroom observation. It may not include everything that might be desirable: it includes only minimal discussion of and a single resource for peer review of faculty assessment of their students' work, too little on peer review of teaching as scholarly activity, and limited links between teaching and course portfolios as vehicles documenting scholarship of teaching. But she either has introduced or gone well beyond what others have contributed in these areas. Chism herself and others, no doubt, will further the discussion of these topics and contribute even more powerful resources in the future. This book deserves thoughtful reading and consideration by faculty and administrators, and the resources in it should prove valuable as peer review of teaching becomes more commonplace at colleges and universities.

References

Anderson, R. S., & Speck, B. W. (1998, Summer). *Changing the way we grade student performance: Classroom assessment and the new learning paradigm.* New Directions for Teaching and Learning No. 74. San Francisco: Jossey-Bass.

Barr, R. B., & Tagg, J. (1995, November/December). From teaching to learning—A new paradigm for undergraduate education. *Change*, 13–24.

Batista, E. E. (1976). The place of colleague evaluation in the appraisal of college teaching: A review of the literature. *Research in Higher Education, 4,* 257–271.

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate.* Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Centra, J. A. (1986, April). *Colleague evaluation: The critical link.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Documentation Reproduction Service No. 275 722)

Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness.* San Francisco: Jossey-Bass.

Cohen, P. A., & McKeachie, W. J. (1980). The role of colleagues in the evaluation of college teaching. *Improving college and university teaching, 28*(4), 147–154.

Cross, K. P. (1991, Fall). College teaching: What do we know about it? *Innovative Higher Education, 16*(1), 7–25.

Keig, L., & Waggoner, M. D. (1994). *Collaborative peer review: The role of faculty in improving college teaching.* ASHE-ERIC Higher Education Report No. 2. Washington, DC: The George Washington University, School of Education and Human Development.

Keig, L. W., & Waggoner, M. D. (1995). Peer review of teaching: Improving college instruction through formative assessment. *Journal on Excellence in College Teaching, 6*(3), 51–93.

McKeachie, W. J. (1994). *Teaching tips: Strategies, research, and theory for college and university teachers* (9th ed.). Lexington, MA: Heath.

Scriven, M. S. (1980). *The evaluation of college teaching.* Syracuse, NY: National Council of States on Inservice Education. (ERIC Document Reproduction Service No. 203 729)

Scriven, M. S. (1985). New frontiers of evaluation. *Evaluation Practices, 7*(1), 7–44.

Scriven, M. S. (1987). Validity in personnel evaluation. *Journal of Personnel Evaluation in Education, 1,* 9–23.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Shulman, L. S. (1999). Course anatomy: The dissemination and analysis of knowledge through teaching. In P. Hutchings (Ed.), *The course portfolio: How faculty can examine their teaching to* advance *practice and improve student learning* (pp. 5–12). Washington, DC: American Association for Higher Education.

Soderberg, L. O. (1986, March). A credible model: Evaluating classroom teaching in higher education. *Instructional Evaluation, 8,* 13–27.

Book Reviews

Enhancing Student Learning: Setting the Campus Context, edited by Frances K. Stage, Lemuel W. Watson, and Melvin Terrell. University Press of America, 1999. 145 pp. $24.00 (paper).

PETER M. MAGOLDA, Miami University

Stage, Watson, and Terrell's primary aim in *Enhancing Student Learning* is to encourage the higher education community—in particular student affairs—to refocus its attention on student learning. In this eight-chapter book, the editors and six additional contributors first make the case that student learning is important, citing numerous instrumental benefits. Next they introduce relevant theories and then provide case studies that demonstrate how these theories could be applied to many higher education settings. The book concludes with a chapter that provides a primer for assessing student learning and a chapter that synthesizes the contributors' main points and speculates on the future. In short, this book is chock-full of reminders and helpful hints that would advance the authors' aim of enhancing student learning. The book does not target readers who desire ground-breaking theoretical perspectives based on original research.

Watson and Stage's introductory chapter encourages readers to think holistically when exploring the complex terrain of student learning. Watson's conceptual model of learning, involvement, and gains is the theoretical framework introduced to accomplish this aim. The model posits that one must concurrently focus on input (e.g., influences of secondary education) process (e.g., the collegiate environment), and output (e.g., educational gains) to influence student learning. Collectively, the model and the case study that follows remind readers of a seldom-contested assertion in student affairs—student learning is shaped by many factors, some of which extend beyond the scope of the classroom (i.e., the traditional venue for "learning").

The "Theories of Learning for College Students" chapter written by Frances Stage and Patricia Muller begins by discussing the usefulness and importance of theory. An eclectic array of theories—self-efficacy, constructivist learning, multiple intelligences, learning styles, and conscientization—follow. These diverse theoretical overviews (i.e., that favor breadth rather than depth) and the accompanying case study examples target student affairs professionals and their work contexts.

Watson and Terrell in their "Cultural Differences in Student Learning and Development" chapter promote a culture-oriented education, whereby educators accommodate the preferred learning styles of culturally diverse students. Watson's learning, involvement, and educational gains model is again used to clarify the notion of culture in learning. Discussions about subcultures and differences—in particular how some subcultures have been historically disadvantaged and marginalized on college campuses—dominate this chapter. A more explicit discussion of some of the sociopolitical issues that are seldom explicitly discussed would have been a welcome addition.

The authors of chapters 4 to 6 apply the theories explained in the opening chapters to diverse contexts. In the "Learning and Development from Theory to Practice," chapter, Cuyjet and Lewis Newman provide familiar examples of how theory can be used to deliver high quality services and programs to students. Four key issues—facilitating the needs of new students, developing community, developing value, and instilling a service mindset— central to the student learning agenda are detailed. De la Teja and Kramer's "Student Affairs and Learning in the College Community" chapter focuses on student learning in community colleges. This example crystallizes for readers the interrelationship between context and student learning. Their analyses highlight the diverse and overlooked learning needs of diverse student populations and marginalized educational settings. In the "Service Learning: Exemplifying the Connections between Theory and Practice" chapter, Muller and Stage apply four learning theories discussed earlier in the text to service-learning programs. These discussions reinforce a reoccurring theme in the book and in student affairs—student affairs professionals should play a central role in enhancing student learning.

The assessment chapter, written by Upcraft, begins by clarifying the distinction between assessment and evaluation. Later, Upcraft offers step-by-step tips for novices desiring to measure learning quantitatively and qualitatively. The recommended procedures appear deceptively linear and problem-free. In the concluding chapter, "Setting a New Context for Student Learning," Stage and Watson revisit the contributors' main arguments and briefly speculate on a future agenda for higher education.

This book has notable strengths. The authors' arguments are clearly and efficiently presented using a practitioner discourse that is rooted in published research findings and practical work experiences. The case studies illustrate numerous "how-to-get-started" suggestions for student affairs staffs who want to work in harmony with faculty and other academic affairs staff.

There are two notable shortcomings in this text. The first has to do with the authors not always "practicing what they preach." For example, Stage and Muller introduce four theoretical perspectives, some of which are not commonly associated with the student learning imperative; these diverse perspectives are assets. Unfortunately, the presentation of these perspectives violates one of the primary themes of the book—holistic learning. As a reader, I wanted to know, for example, how Conscientization theory, rooted in liberation theology, interconnects with the multiple intelligences discussion that is rooted in cognitive science. The authors' interdisciplinary meaning making is absent, creating the perception that the choices were haphazard. This perception is reinforced when some of the primary literature linking intellectual development with student learning (e.g., see King's [1996] synthesis of research on student cognition and learning) is barely mentioned.

This absence of holistic understanding is also evident when one looks across the eight chapters. Each chapter contributes to enhancing the readers' understanding of their student learning agenda, but a sense of disconnection prevails.

For example, in the second half of the book the discussion progresses from a dialogue about community colleges, to service learning, concluding with an assessment discussion. These three topics are important and could interrelate, but the connections are seldom made. The sense of wholeness, which is a professed core value of the editors, is too often absent.

A second shortcoming centers on the editors' goal of advancing "new ways of looking at old problems" (p. 2). The literature the editors cite in the introductory chapter suggests that their student learning agenda has been around for at least thirty years. Explicitly discussing reasons why academia has been reluctant to get on the "student learning bandwagon" would have been a welcome addition to the text, especially because many of the contributors' suggestions have been previously recommended. Specifically addressing past impediments is an important prerequisite to advancing "new ways of looking at old problems."

Stage, Watson, and Terrell suggest that the intended audience is broad, including student affairs professionals, faculty, and preparatory faculty of student affairs professionals. These intended audiences would be well served by the book if they were unfamiliar with why student learning is central to the mission of student affairs, in particular, and higher education, in general, or if readers were planning a student learning initiative for the first time.

References

King, P. M. (1996). Student cognition and learning. In S. Komives and D. B. Woodard (Eds.), *Student services: A handbook for the profession* (pp. 218–243). San Francisco: Jossey-Bass.

Prioritizing Academic Programs and Services: Reallocating Resources to Achieve Strategic Balance, by Robert C. Dickeson. Foreword by Stanley O. Ikenberry.
San Francisco: Jossey-Bass, 1999. 192 pp. $34.95.

DON G. CREAMER, Virginia Polytechnic Institute and State University

Stanley Ikenberry says that there is no more difficult or more important job in higher education than clarifying purposes and setting priorities for colleges and universities. Accomplishing this task is vital because institutions do not have inexhaustible resources, and the link between academic quality and financial health of institutions is clear. Robert Dickeson offers a road map for setting and shaping these priorities.

Dickeson gained his expertise in the trenches, serving as a teacher, an academic administrator, a university president, and a private organization consultant with leaders of colleges and universities. His mission in this book is to apply what he learned from his experiences to help institutions achieve a tighter focus of academic programs and services and to make better use of their resources.

He accomplishes this purpose by providing a "step-by-step guide for confronting the many dimensions of academic program prioritization" (p. xix).

The author begins by recounting internal and external pressures for reform in higher education and states seven premises for reform. Among the premises are the centrality of academic programs to the collegiate institution, the absence of judgments of worth of an ever increasing number of academic programs, and the failure of traditional across-the-board cuts to adequately address institutional needs to reallocate resources. He also argues that adding new academic programs under current conditions results in diminution of resources for existing programs and that there is too much of what he calls "runaway specialization" in colleges and universities. Curriculum creep leads to program creep that leads to mission creep in his view. In the end, Dickeson contends that "the price for academic program bloat for all is impoverishment of each" (p. 17).

Dickeson's plan recognizes the vital requirement for leadership to accomplish program prioritization and reallocation of resources. He addresses specifically the roles of the president, the board, and the provost and points out that leadership and courage are required. He asks these leaders a series of questions as a form of "gut check" about whether the institution is prepared to undergo the process of prioritization. If they are, then Dickeson calls for a reaffirmation of the institutional mission. In this process, he seeks clarity, but he recognizes some special issues such as political considerations, accreditation, and the reality that purposes change over time. A major reason that clarity is essential to achieve consensus about mission is that "centrality to mission" is a major criterion for judging program worth in Dickeson's plan.

Defining programs is problematic, even in Dickeson's scheme, but he offers an operational definition that seems logical for his purposes. A program is "any activity or collection of activities of the institution that consumes resources" (p. 44). The requirement is to identify programs, not departments, and to identify programs in their most elemental form. It is at this basic level that the judgments about program quality and centrality must be made. For the record, Dickeson seeks to evaluate student affairs programs just as he does academic programs, but frankly, does not address this process thoroughly.

Dickeson suggests using 10 indicators, some quantitative and some qualitative, against which to judge program quality and centrality to mission:

1. History, development, and expectations of the program;
2. External demand for the program;
3. Internal demand for the program;
4. Quality of program inputs and processes;
5. Quality of program outcomes;
6. Size, scope, and productivity of the program;
7. Revenue and other resources generated by the program;
8. Costs and other expenses associated with the program;
9. Impact, justification, and overall essentiality of the program; and
10. Opportunity analysis of the program.

In keeping with his step-by-step approach, Dickeson suggests precisely what to do and what data and in what order data should be collected. He suggests scaling methods for the 10 indicators and precisely the number of points to be assigned to each. After data are collected, he calls for "levels of judgment"; that is, that departments should rank the data first, then directors or deans, vice presidents, and finally, the president. After this ranking process by levels, further ranking is suggested for all programs by categories to yield five levels of possible action. The upper 20% may be candidates for enrichment, the next 20% might be retained for higher level of support, the next 20% retained at neutral level of support, the next 20% retained at a lower level of support, and the lowest 20% become candidates for reduction, phasing out, or consolidation. In Dickeson's scheme, each category must contain exactly 20% of the institutional programs.

Dickeson anticipates questions that will be raised about the prioritization process and suggests some process remedies, then addresses decisions about enrichment or expansion of current programs and adding new programs. He confronts issues associated with reduction of programs, elimination of programs, and consolidation or restructuring of programs. He acknowledges legal and policy implications such as reduction in force (RIF) policies and tenure issues. He discusses accreditation issues and humane dimensions of making serious program changes.

In conclusion, Dickeson summarizes the external and the internal forces that affect program prioritization and suggests that they be used in a force-field analysis to help institutions achieve a "strategic balance" in teaching, research, and service; purposes; fiscal expectations; congruence; and affordability and accessibility. Strategic balance also is sought in stability and flexibility; harmonizing institutional interest and public interests; respecting tradition and readying for the future; reconciling competing expectations; integrating liberal arts and career preparation; planning top down and bottom up; and delineating authority and responsibility. An ambitious agenda indeed!

The step-by-step plan for program prioritization offered in this book may be particularly useful for college and university administrators and for members of boards of trustees. The book does not make a unique contribution, however, except in its progressive and step-by-step approach to program restructuring. Calls for reform of program priorities are common in higher education literature. David Leslie and E. K. Fretwell, Jr., in *Wise Moves in Hard Times*, for example, devote a section to searching for solutions just as Dickeson described. In *Restructuring Higher Education*, Terrence MacTaggart discusses contexts for change, case studies, and lessons learned from restructuring efforts that are driven by the same forces that Dickeson identifies. Likewise, Ami Zusman specifically includes institutional retrenchment and reallocation in a discussion of "Issues in Higher Education" (found in Philip Altbach, Robert Berdahl, and Patricia Gumport, *American Higher Education in the Twenty-first Century*).

The book has its shortcomings. It is never clear, for example, whether Dickeson is talking about all academic programs within an institution or whether he is talking only about undergraduate programs. He tries to differentiate between major institution types in his plan, but this is not accomplished very successfully, because he provides only surface distinctions and does not address the use of his plan by type of institution. His call for comprehensive reform leaves one to wonder whether his arguments for reform really suggest that department heads, deans, vice presidents, and provosts are not doing their jobs very well and that if they were, the kind of pervasive tactic Dickeson supports might not be necessary. Further, one wonders why such a extensive tactic as Dickeson advocates would not be better accomplished in concert with some other compelling need to conduct a self-study, such as during regional accreditation reaffirmation. Dickeson never explains why this approach is best conducted as a stand-alone initiative.

Overall, the book is worthwhile for institutional leaders trying to achieve tighter focus of academic programs and services. Because this is a general interest of these leaders, the book should have a widespread appeal.

References

Altbach, P. G., Berdahl, R. O., & Gumport, P. J. (Eds.). (1999). *American higher education in the twenty-first century: Social, political, and economic challenges.* Baltimore, MD: The Johns Hopkins University Press.

Leslie, D. W., & Fretwell, E. K., Jr. (1996). *Wise moves in hard times: Creating and managing resilient colleges and universities.* San Francisco: Jossey-Bass.

MacTaggart, T. J., & Associates. (1996). *Restructuring higher education: What works and what doesn't in reorganizing governing systems.* San Francisco: Jossey-Bass.

Zusman, A. (1999). *Issues facing higher education in the twenty-first century.* In P. G. Altbach, R. O. Berdahl, & P. J. Gumport (Eds.), *American higher education in the twenty-first century: Social, political, and economic challenges* (pp. 109–148). Baltimore, MD: The Johns Hopkins University Press.

Shattering the Myths: Women in Academe, by
Judith Glazer-Raymo. Baltimore, MD: The Johns
Hopkins University Press, 1999. 208+ pp. $38.00

FLORENCE A. HAMRICK, Iowa State University

In *Shattering the Myths,* the author uses feminist perspectives to analyze gender representation in higher education and the persistence of gender-related disparities. She also analyzes several prevailing assumptions and practices in which disparities are anchored that serve to hinder women's progress within higher education. These include seemingly gender-neutral practices such as faculty employment tracks, assignment of merit criteria, and specification of accountability measures. For the most part, these are the myths that Glazer-Raymo seeks to expose and, in turn, shatter through presenting evidence that "gender neutrality is a fiction in the professoriate and in academic administration" (p. 24). She then provides and discusses suggestions and recommendations for promising strategies.

In Chapter 1, "The Personal and the Professional: Becoming a Feminist," Glazer-Raymo recalls more than thirty years of her own life as an educator, administrator, and feminist in light of evolving social and political developments like Title IX, affirmative action policies, and the expansion of higher education. Major trends and issues identified in the first chapter are woven through the remainder of the book. Some readers may be impatient with the biographical technique in the first chapter, because the narrative clearly draws the author's life (Reinharz, 1992) into a scholarly examination of public policy issues and problems. However, this technique effectively reinforces the message that lives unfold against sociopolitical backdrops that shape assumptions and careers, and these backdrops and assumptions must be well understood if change is to be possible.

In Chapters 2 through 5, Glazer-Raymo examines the academic labor market; tenure, promotion, and salaries; professional women (e.g., lawyers, dentists); and women administrators. Some common problems are echoed, such as women's underrepresentation in male-dominated and higher paying disciplines, affecting not only women's prestige but bargaining power. For example, she demonstrates how women in male-dominated professions enter not a "free market" of ideas or efforts, but enter instead a complex network of inherited structures that function to advantage traditional insiders whose values and privileges the structures manifest. Glazer-Raymo discusses the subtle forms of harassment and segregation that are evidence of general reluctance toward inclusiveness, equity, and valuing women's scholarship. However, the news is not uniformly positive with respect to legal redress, based on Glazer-Raymo's summary of the mixed success with respect to individual court cases.

These middle chapters synthesize evidence from case law and legislation, employment figures, and enrollment statistics. In her final analysis, echoing Tierney and Bensimon's (1996) work, access to education—even advanced education—and entry into (academic) work are not the sites of the most intractable problems faced by women. Advancement, equitable remuneration, and valuing the contributions of people who largely remain cultural outsiders are the persistent problems. The reasons for these disparities are complex, and problem sources are not easily identified or essentialized. However, Glazer-Raymo's book excels at identifying issues, contextualizing them within social and political climates that are continually evolving and substantiating explanations for their persistence.

In the final chapters on implementing change and conclusions, Glazer-Raymo discusses two potential strategies for change—campus commissions and feminist pedagogy. Campus commissions focus attention on local evidence of disparities and, more recently, defend existing operations and programs in the current climate of feminist backlash. Several examples of effective commissions were cited. Glazer-Raymo's promotion of localized study and action through commissions as an effective strategy subsequently has been reinforced by the recent publicity surrounding documentation of gender discrimination at MIT's School of Science (Wilson, 1999) and by plans for institutional redress. Time will tell whether local commission strategies are effective for groups on other campuses who are now engaged in similar efforts. Following the MIT example, what may perhaps emerge are derivative commissions that focus not on an entire campus with its broad collection of disciplines and fields of study but instead on colleges or subsets of related disciplines with somewhat more shared notions of discrimination indicators (e.g., laboratory space among scientists). Such a trend would also be consistent with Glazer-Raymo's earlier analysis of feminism proceeding from its early emphasis on broadly-focused, collective social action to its later emergence in subgroups—albeit not always respected subgroups—within individual professional or disciplinary associations.

With respect to the second strategy of feminist pedagogy, Glazer-Raymo claims: "Engendering pedagogical theory and practice and building on knowledge gained through two decades of feminist pedagogy provides practical strategies for changing classroom and campus cultures" (p. 195). There seems little doubt that feminist pedagogy (features of which are also echoed in adult education literature) serve to enrich the learning environment for students. It is less clear that improved pedagogical practices will lead to campus cultural or organizational transformation—particularly a transformation that will benefit women academics who enlist these pedagogies. Despite more widespread calls for effective teaching (e.g., Wingspread Group, 1993) and attempts to reinterpret teaching as a form of scholarship (e.g., Boyer, 1990), there is less persuasive evidence that the research-teaching-service scaffold has been successfully toppled at individual campuses. This is especially the case for prestigious institutions where stature is clearly grounded in research and grant activity. Unfortunately, until a shift in thought and action becomes apparent, emphasizing teaching and feminist pedagogy may serve instead as reinforcement of the notion of women performing the women's work of academe (Park, 1996).

In summary, *Shattering the Myths* is an extremely important book. Glazer-Raymo highlights evolving social, political, and institutional forces surrounding gender issues and women's representation as well as the myths shaping perceptions of gender equity in academe. She clearly documents evidence of persistent disparities; reveals multiple, reinforcing sources of discrimination; and offers strategies to address discrimination. Perhaps most importantly, Glazer-Raymo demonstrates that although gender discrimination has evolved and now appears in different forms, it nonetheless remains a characteristic feature of academe, and women academics continue to pay a disproportionate price.

References

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate.* Lawrenceville, NJ: Princeton University Press.

Park, S. M. (1996). Research, teaching, and service: Why shouldn't women's work count? *Journal of Higher Education, 67,* 46–84.

Reinharz, S. (1992). *Feminist methods in social research.* New York: Oxford University Press.

Tierney, W. G., & Bensimon, E. M. (1996). *Promotion and tenure: Community and socialization in academe.* Albany, NY: State University of New York Press.

Wilson, R. (1999). An MIT professor's suspicion of bias leads to a new movement for academic women. *Chronicle of Higher Education, 46*(15), pp. A16–A18.

Wingspread Group on Higher Education. (1993). *An American imperative: Higher expectations for higher education.* Racine, WI: Johnson Foundation.

Instructions to Contributors

Manuscripts should be mailed in triplicate to Leonard L. Baird, Editor, *The Journal of Higher Education,* Ohio State University Press, 1070 Carmack Road, Columbus, OH 43210-1002.

Style. *The Journal of Higher Education* has adopted as its official guide the *Publication Manual of the American Psychological Association,* fourth edition, and all manuscripts should be brought into conformity with this guide before they are submitted. Papers should be typed, double-spaced, on white 8½ × 11-inch paper, with wide margins. Upon acceptance, contributors will be asked to supply a computer file on an IBM-compatible standard disk in WordPerfect. Authors who use a different system should prepare an ASCII text file. An abstract of fifty words or less, summarizing the main points of the article, should accompany the manuscript. Because the journal's readers represent a variety of professional interests, it is recommended that any statistical material be presented as briefly and simply as possible. Although each paper submitted should deal with the methodology employed in addressing the subject in sufficient detail to place the data within the proper methodological setting, the editors of the journal are not primarily interested in papers setting forth practices of research methods (an acquaintance with fundamental procedures of scholarly analysis being assumed on the part of the reader) except for those papers that develop innovative methodological approaches.

Illustrations submitted with the final draft must be of professional quality, and executed on white paper or vellum, in black ink, with clear, medium weight, black lines and figures. Typewritten lettering should not appear in illustrations. Figures should be provided **at size**—no larger than 4½ × 7 inches (full page) and preferably no larger that 4½ × 3½ inches (half page) and **printed camera-ready at a minimum of 600 dpi.** They should be numbered consecutively, and the number and author's name should be penciled lightly on the back of each. All illustrations must have captions, which should not appear on the artwork but should be typed, double-spaced, on a sheet at the end of the manuscript. If there is any potential for doubt, the word *top* should be written on the back of the illustration.

Authors should employ a Reference List format to list bibliographic data. Endnotes should be reserved for supplementary comment and typed on a separate page at the end of the manuscript.

Manuscript Length. Manuscripts averaging 25–30 pages of double-spaced typescript are preferred, but the editors will consider longer papers on topics requiring a fuller treatment.

Review Process. Those unsolicited manuscripts that are refereed are reviewed blind. Authors are thus requested to submit their name, professional position, and institution on a removable cover sheet. They should also mask any items of self-reference where they appear. Authors must not submit the manuscript of any article that is still under consideration by another publisher.

Editorial Reaction. Papers will not be returned to authors if they fail to meet by a wide margin the basic criteria for selection. Otherwise, authors may expect to receive some notification within three months. If an article is accepted, it will usually appear in print within twelve months after acceptance. If an article that has been subjected to a full review is rejected, the opinions of the referees will be transmitted to the author.

Criteria for Selection. Papers are evaluated on the following points: Form: writing style and readability, logical development, appropriate length, appropriateness of author's stated objectives to treatment as defined below. Content: significance to *JHE* readers. Additional criteria are based upon the following manuscript orientations: as a research paper, as a technical paper, as a professional practice paper, as a literature review, and as a policy paper. It should be emphasized that the editors respond most favorably to manuscripts that evidence both a freshness of vision and a vitality that may be informed by, but certainly go beyond, methodological qualities, and that are in congruence with our publishing goals and directions. The most effective approach in learning about our interests is to read previous issues of the journal. We expect that authors, the journal, and the field will develop through the publication process.

Marvin W. Peterson
Marne K. Einarson

What Are Colleges Doing About Student Assessment?

Does It Make a Difference?

Introduction

An increasing number of colleges and universities have engaged in some form of student assessment activity over the past decade (El-Khawas, 1988, 1990, 1995). Administrators and faculty have invested considerable time and effort in promoting, supporting, and implementing these assessment efforts. Yet concerns have been raised that assessment activities are difficult to mount successfully (Ewell, 1988a; Gray & Banta, 1997) and seldom produce discernible impacts on students' or institutions' performance (Astin, 1991; Ratcliff & Associates, 1995).

Student assessment, particularly scholarship taking an institutional perspective, is an emerging arena of study in higher education. A rich prescriptive literature exists. Scholars and practitioners have produced many books and monographs that offer guidelines for how institutions might best approach and support their student assessment efforts (cf. Banta & Associates, 1993; Banta, Lund, Black, & Oblander, 1996; Rossman & El-Khawas, 1987; Sims, 1992). Less available is nationally representative empirical evidence concerning how institutions have conducted student assessment and to what effect. Studies have collected descriptive data regarding the content and methods of institutions' assessment approaches (Cowart, 1990; Ory & Parker, 1989; Steele & Lutz,

Marvin W. Peterson is a professor at the Center for the Study of Higher and Postsecondary Education, The University of Michigan. Marne K. Einarson was a senior research associate at the School of Education, The University of Michigan, when this article was written; she is now a senior research and planning associate at the Office of Institutional Research and Planning, Cornell University.

The Journal of Higher Education, Vol. 72, No. 6 (November/December 2001)
Copyright © 2001 by The Ohio State University

1995). But there has been little systematic examination of organizational and administrative patterns at the institutional level developed to support student assessment efforts (Ewell, 1997; Johnson, Prus, Andersen, & El-Khawas, 1991) or how institutions have used and been affected by assessment information (Banta et al., 1996; Ewell, 1988b; Gray & Banta, 1997).

The purpose of our study was to extend current understanding of how postsecondary institutions have approached, supported, and promoted undergraduate student assessment, and the institutional uses and impacts that have been realized from these assessment efforts. In addition, we were interested in examining the congruence between institutional approaches to student assessment found in the prescriptive literature and actual institutional practices.

Review of the Related Literature

We began our research by conducting an extensive review of the literature on student assessment in postsecondary institutions. Details of this literature review and its results are available elsewhere (see Peterson & Einarson, 2000; Peterson, Einarson, Trice, & Nichols, 1997). For this study, we focused on the following conceptual domains: (1) the relationship of institutional context to student assessment;(2) institutional approaches to student assessment;(3) organizational and administrative support for student assessment;(4) assessment management policies and practices; and (5) institutional uses and impacts of student assessment information. Our review considered prescriptive literature offering recommendations for institutions embarking on student assessment and empirical literature providing evidence of institutional practices and consequences with respect to student assessment. Findings from the literature review are summarized below.

Institutional Context

The literature suggests that variations in the student assessment practices observed across postsecondary institutions are related to differences in broad characteristics of institutional context. Differences by institutional type have been observed in the content and methods of student assessment approaches (Johnson, Dasher-Alston, Ratteray, & Kait, 1991; Steele & Lutz, 1995; Steele, Malone & Lutz, 1997). Fewer studies have found differences in forms of organizational and administrative support for student assessment by institutional type (Patton, Dasher-Alston, Ratteray, & Kait, 1996), control (Johnson et al., 1991) and size (Woodard, Hyman, von Destinon, & Jamison, 1991). These findings suggest that patterns of assessment management policies and practices and institutional uses and impacts from assessment may also differ by institutional type, control, and size. No such comparative research was found.

Institutional Approaches to Student Assessment

Institutional approach to student assessment refers to the content and technical aspects of student assessment. Three dimensions of institutions' approaches to student assessment were relevant to this study: the content or type of student data collected, methods employed to collect assessment data, and the analyses conducted with assessment data.

Institutions may collect data concerning students' cognitive (e.g., higher-order cognitive processes, subject-matter knowledge), affective (e.g., values, attitudes) and behavioral (e.g., course completion, hours spent studying) performance or development (cf. Alexander & Stark, 1986; Astin, 1991; Bowen, 1977; Ewell, 1984; Lenning, Lee, Micek, & Service, 1977). Second, institutions may employ various methods to collect assessment data. Choices include comprehensive examinations or inventories; integrative or performance-based methods such as capstone projects, demonstrations, or portfolios; surveys or interviews; or the collection of institutional data such as enrollment or transcript data (Ewell, 1987c; Fong, 1988; Johnson, McCormick, Prus, & Rogers, 1993). Third, assessment approaches vary in the nature of analyses conducted with collected data. Institutions may analyze data at various levels of aggregation such as individual students or student subgroups, courses, academic programs or departments, schools or colleges, or the institution as a whole (Alexander & Stark, 1986; Astin, 1991; Ewell, 1984, 1988b). Approaches also vary in the complexity of analyses conducted. Analysis options range from producing descriptive summaries of student data to conducting relational studies of students' performance to aspects of their educational experiences such as course-taking patterns, exposure to different teaching methods, and interactions with faculty (Astin, 1991; Ewell, 1988b).

Scholars assert that a positive relationship exists between institutional uses and impacts of assessment and the comprehensiveness of an institution's assessment approach—that is, the extent to which an institution collects data on various domains of student functioning (Astin, 1991; Ewell, 1988b; Johnson et al., 1993), uses a variety of assessment methods (Ewell, 1984, 1988b; Halpern, 1987; Jacobi, Astin, & Ayala, 1987; Terenzini, 1989), analyzes data at several levels of aggregation (Astin, 1991; Ewell, 1988b), and studies the relationship between student performance and institutional experiences (Ewell, 1988b, 1991; Johnson et al., 1993).

Extant research shows institutions have adopted circumscribed approaches to student assessment. Although variance in content emphasis has been observed among types of institutions, most have focused on assessing cognitive rather than behavioral or affective domains of student development (Cowart, 1990; Gill, 1993; Johnson et al., 1991; Patton et al., 1996; Steele & Lutz, 1995; Steele et al., 1997). There is conflicting evidence regarding the assessment methods used by institutions. Some studies observed an increase in the tendency of institutions to use complex measures such as portfolio assessment (El-Khawas, 1992, 1995), whereas others found that institutions were most likely to rely on easily collected and quantifiable measures such as course completion and grade data (Cowart, 1990; Gill, 1993; Patton et al., 1996; Steele & Lutz, 1995). No research was located that examined institutional choices regarding the level of aggregation of data analysis. There is limited evidence of institutions conducting relational studies of student performance and institutional experiences (Gill, 1993; Patton et al., 1996; Steele et al., 1997).

Organizational and Administrative Support for Assessment

Three dimensions of organizational and administrative support have been discussed in the literature as important influences on the effectiveness of institutions' student assessment efforts: institutional support strategy, leadership and governance patterns, and evaluation of student assessment processes.

Institutional support strategy refers to an institution's choices about the overall purpose or functions of its assessment efforts. Scholars propose that assessment strategies that emphasize internal purposes such as institutional improvement rather than external purposes such as meeting state-level or accreditation requirements will foster internal support for assessment and enhance the likelihood of garnering positive impacts from assessment (Aper, Cuver, & Hinkle, 1990; Braskamp, 1991; Ewell, 1987a). An institution's academic mission reflects strategic choices regarding educational goals and activities to be pursued. Whether or not an institution's academic mission prioritizes undergraduate teaching and learning(Banta & Associates, 1993; Hutchings & Marchese, 1990) and student assessment (Duvall, 1994) as valued institutional activities or specifies intended educational outcomes (Braskamp, 1991) may be predictive of assessment support and impacts.

Supportive leadership on the part of senior administrators (Banta et al., 1996; Ewell, 1988a; Rossman & El-Khawas, 1987) and faculty (Banta & Associates, 1993; Sell, 1989; Young & Knight, 1993) is thought to play a critical role in promoting an institution's student assessment efforts. The nature of governance processes for student assessment is also expected to influence internal support for assessment and the likelihood of achieving assessment impacts. On the one hand, a centralized approach as reflected in the existence of a formal policy for student assessment or creation of a central coordinating office may do much to symbolize an institution's commitment to assessment (Ewell, 1984; Thomas, 1991). More often, scholars have advocated a decentralized governance approach involving a broad range of internal participants, particularly faculty, as a means of enhancing the acceptance and use of assessment results (Astin, 1991; Banta et al., 1996; Eisenman, 1991; Ewell, 1984; Terenzini, 1989).

Finally, student assessment is expected to contribute most to the improvement of student and institutional performance when it is approached as a constantly evolving process rather than an episodic activity (Wolff, 1992). Accordingly, scholars have characterized evaluation of the assessment process as an important means by which institutions can enhance the institutional impact of their student assessment efforts (Dennison & Bunda, 1989; Ewell, 1988b; Sims, 1992).

Institutions have reported engaging in student assessment activities for a variety of internal and external purposes (El-Khawas, 1995). Descriptive data suggest that assessment support strategies that give equal or greater weight to internal improvement purposes than to external accountability purposes have developed more comprehensive student assessment approaches (Hyman, Beeler & Benedict, 1994; Johnson et al., 1991; Muffo, 1992). There is limited evidence that an increasing number of institutions have developed institution-wide student assessment plans (Gill, 1993; Patton et al., 1996) and that the centralization of decision making varies across the phases of planning, implementing, and using information from student assessment (Johnson et al., 1991; Patton et al., 1996). Results from one study suggest that few institutions have evaluated their student assessment approach (Patton et al., 1996).

Assessment Management Policies and Practices

Scholars contend the extent to which institutions devise specific policies and practices to support their student assessment activities and integrate them with other institutional processes is a powerful means of promoting internal support for and institutional impacts from assessment (Ewell, 1988a; Sell, 1989). Institutions have been encouraged to link internal resource allocation processes to their assessment efforts (Ewell, 1984, 1987a, 1987b, 1987c, 1988a; Thomas, 1991); to develop comprehensive computerized student assessment information systems to manage collected data (Astin, 1987; Ewell, 1984, 1988a; Terenzini, 1989);

to regularly communicate student assessment purposes, activities, and results to internal and external constituencies (Ewell, 1984; Mentkowski, 1991; Thomas, 1991); to devise policies encouraging student participation in assessment activities (Duvall, 1994; Erwin, 1991; Loacker & Mentkowski, 1993); to provide professional development on assessment-related topics (Ewell, 1988b; Gentemann, Fletcher, & Potter, 1994; Young & Knight, 1993); and to integrate student assessment with processes for academic planning and review (Barak & Sweeney, 1995; Chaffe-Stengel, 1992; Ewell, 1984, 1988a, 1997). Conflicting opinions have been offered as to whether institutions should link policies concerning faculty evaluation and rewards to student assessment involvement or results (Ewell, 1984; Halpern, 1987; Ryan, 1993).

A few studies have provided descriptive data on the extent to which institutions have implemented policies and practices regarding communication of assessment information (Patton et al., 1996), student assessment information systems (Astin, 1987; Gill, 1993), faculty development (Steele & Lutz, 1995), and program review (Barak & Sweeney, 1995) to support their assessment efforts. In one relational study, the breadth of the intended audience for assessment reports was predictive of achieving positive outcomes from institutions' student assessment projects but resource allocation practices were not (California State University Institute for Teaching and Learning [CSUITL], 1993). Despite the reputed importance of this dimension of student assessment support, it has received limited attention in empirical studies.

Uses and Impacts of Student Assessment Information

Scholars suggest that student assessment information can be used in many areas of institutional planning and decision making, including institutional mission and goals (Banta et al., 1996; Ewell, 1984; Jacobi et al., 1987); academic programs (Ewell, 1987a, 1987b, 1988a, 1997; Gentemann et al., 1994); student support services (Banta, 1985; Williford & Moden, 1993; Young & Knight, 1993); academic resource allocation (Ewell 1984, 1987a, 1987b, 1987c, 1988b; Thomas, 1991); and faculty evaluation and rewards (Ewell, 1984, 1988b; Jones & Ewell, 1993).

The ultimate measure of the effectiveness of an institution's student assessment efforts is the extent to which these efforts contribute to improvements in student and institutional performance (American Association for Higher Education [AAHE], 1992; Banta & Associates, 1993; Ewell, 1987b, 1988b, 1997). Clearly, enhancing students' academic performance is an important potential institutional benefit of engaging in student assessment. In addition, assessment may have positive impacts on faculty behavior and attitudes and on institutions' relationships with the external environment (Banta & Associates, 1993; Banta et al., 1996; Ewell, 1997).

The bulk of evidence concerning institutional uses of assessment information comes from descriptions of single institutions' assessment efforts (cf. Banta et al, 1996; Banta & Associates, 1993). Limited evidence from multi-institutional research reveals that institutions have most often used assessment information in decisions about program review (Barak & Sweeney, 1995) and least often in decisions regarding faculty rewards (Cowart, 1990; Steele & Lutz, 1995).

A few institutions have noted increases in student achievement on standardized examinations (Bowyer, 1996; Krueger & Heisserer, 1987; Magruder & Young, 1993), grade point averages (RiCharde, Olny, & Erwin, 1993), retention (RiCharde et al., 1993; Walleri & Seybert, 1993), and satisfaction with educational experiences (Katz, 1993; Krueger & Heisserer, 1987; Williford & Moden, 1993). However, two multi-institutional studies produced conflicting evidence regarding the impact of student assessment on student grade achievement (CSUITL, 1993; Johnson et al., 1991).

Descriptions of assessment practices undertaken at various institutions suggest that student assessment efforts may stimulate changes in the teaching methods used by faculty, including greater use of teaching approaches that promote student involvement in learning (Banta & Moffett, 1987; Friedlander, Murrell, & MacDougall, 1993; Walleri & Seybert, 1993; Young & Knight, 1993). Reports of individual institutions' assessment experiences also suggest that these efforts may have a positive effect on faculty satisfaction (Katz, 1993; Young & Knight, 1993) and interest in teaching (Friedlander et al., 1993; Hutchings & Marchese, 1990), and may stimulate campus discussions about undergraduate education (Hutchings & Marchese, 1990). Multi-institutional studies found little evidence that student assessment had led to changes in faculty teaching methods (CSUITL, 1993; Cowart, 1990) but have documented more positive faculty attitudes toward assessment activities (CSUITL, 1993), increases in faculty interaction and cooperation (Johnson et al., 1991), and increased sensitivity to the student experience (Hyman et al., 1994) as benefits derived from assessment.

There has been less consideration of the association between institutions' student assessment efforts and relationships with their external environment. The most frequently reported external use of assessment information is responding to state and accreditation reporting requirements (Cowart, 1990; El-Khawas, 1989, 1995; Ory & Parker, 1989). A few institutions have documented a positive relationship between their assessment efforts and allocations of state funding, number and quality of student applicants, and reputation of their institution (Young & Knight, 1993; Williford & Moden, 1993; McClain et al., 1986).

Overall, descriptive data about institutions' assessment approaches are relatively plentiful in the assessment literature. There is comparatively less information available concerning patterns of organizational and administrative support, management policies and practices, uses and impacts of assessment information, and the way these patterns differ by institutional type. The bulk of existing knowledge concerning postsecondary student assessment comes from descriptions of assessment efforts on individual campuses.

Research Questions

The purpose of this study was to provide a current and comprehensive portrait of institutions' approaches to student assessment, specific dimensions of organizational and administrative support for student assessment, assessment management policies and practices, and institutional uses and impacts stemming from assessment. We were interested in examining whether patterns of assessment approaches, organizational and administrative support, management policies and practices, and uses and impacts differed by institutional type. The research questions addressed in this study were:

1. What approaches to student assessment have institutions adopted? How do assessment approaches vary by institutional type?
2. What organizational and administrative support patterns for student assessment have institutions adopted? How do assessment support patterns vary by institutional type?
3. What assessment management policies and practices have institutions adopted? How do assessment management policies and practices vary by institutional type?
4. What uses and impacts from student assessment have institutions reported? How do assessment uses and impacts differ by institutional type?

Conceptual Framework

Based on our literature review, we developed a conceptual framework of institutional support for student assessment comprised of five domains: institutional context, student assessment approaches, organizational and administrative support, assessment management policies and practices, and uses and impacts of student assessment. This framework is displayed in Figure 1.

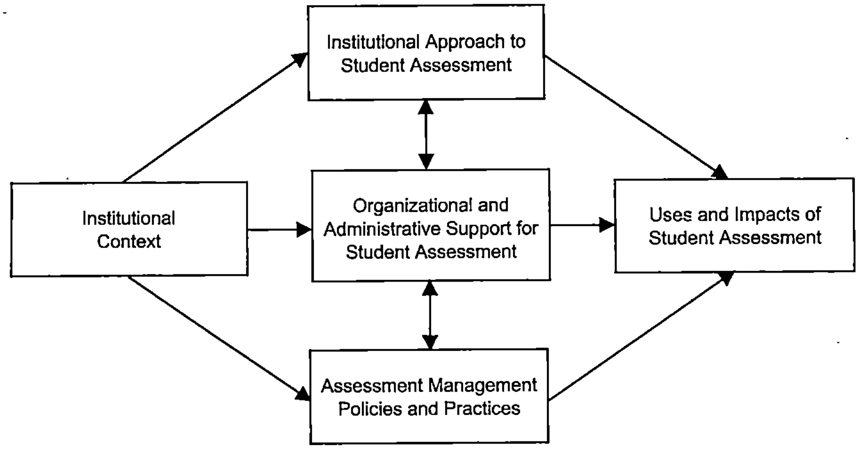
Variables

Variables included in each domain of this framework are listed below. The variable names used in our analyses and tables are italicized. Operational definitions for all variables in the model and reliability estimates where appropriate are provided in Appendix A.

Institutional context. This domain was measured by *institutional type* (associate of arts, baccalaureate, master's, doctoral and research).

Institutional approach to assessment. The content of institutions' assessment approaches was measured by the extent to which they collect data on students' *cognitive*, *affective*, and *postcollege* development or performance. Three variables measured assessment methods—the number of comprehensive tests or inventories used (*number of instruments*), use of integrative or performance-based assessment methods (*student-centered methods*), and use of methods involving external constituencies (*external methods*). Analysis approach was measured by the *number of studies* conducted of student performance and institutional experiences, and the number of levels at which assessment data are aggregated and reported (*number of reports*).

Organizational and administrative support for assessment. Four variables represented institutional strategy for assessment—*mission emphasis* on undergraduate education and assessment, and the importance of internal reasons (*internal purposes*), accreditation requirements (*ac-*



Institutional Approach to Student Assessment
Institutional Context
Organizational and Administrative Support for Student Assessment
Uses and Impacts of Student Assessment
Assessment Management Policies and Practices

FIG. 1. Conceptual Framework of Institutional Support for Student Assessment

creditation purposes), and state requirements (*state purposes*) as purposes for engaging in student assessment. Leadership and governance patterns for assessment were measured by the number of *administrative and governance activities* used to promote assessment, *administrative and faculty support* for assessment, whether the institution has a *formal centralized policy* and uses an *institution-wide planning group* for assessment. Evaluation was measured by whether an institution has formally or informally evaluated its assessment process (*conducted evaluation*).

Assessment management policies and practices. Measures were developed for nine dimensions of assessment policies and practices: whether assessment data are incorporated in *academic planning and review* processes; use of assessment data for *budget decisions*; information system support for assessment (*computer support*), breadth of internal access to assessment information on individual students (*access to information*); breadth of distribution of assessment reports (*distribution of reports*); policies promoting *student involvement* in assessment; *professional development* policies on assessment for faculty and academic administrators; professional development policies on assessment for *student affairs* personnel; and policies linking *faculty evaluation* and rewards to student assessment involvement or results.

Uses and impacts of assessment. Academic decisions measured institutional use of student assessment information in academic decisions such as academic program planning, curriculum, academic support services, and academic resource allocation. *Faculty decisions* measured the extent to which assessment information has influenced institutions' decisions regarding faculty promotion and tenure and salary increases or rewards. Four measures of the impact of student assessment on students were used: *student satisfaction*, *retention and graduation* rates, *grade performance*, and achievement on external examinations (*external achievement*). Four items served as measures of the impact of student assessment on faculty: *campus discussions* of undergraduate education, *faculty satisfaction*, *interest in teaching*, and changes in *teaching methods*. Finally, four items measured the impact of student assessment information on external indicators of institutional performance: *student applications* or acceptance rates, allocation of *state funding*, regional *accreditation evaluation*, and *institutional reputation*.

Method

Sample

The present study is part of a national research effort on student learning and assessment in postsecondary institutions undertaken by Project

Five of the National Center for Postsecondary Improvement. Based on our literature review and pilot testing with chief academic officers in four types of institutions (associate of arts, baccalaureate, doctoral, and research), we developed and refined a survey of "Institutional Support for Student Assessment." The survey is a comprehensive inventory of external influences on student assessment, institutional approaches to student assessment, organizational and administrative policies and practices supporting student assessment, and institutional uses and impacts of student assessment. In January 1998, the survey was mailed to chief academic officers at all postsecondary institutions recognized by the United States Office of Education that offer undergraduate programs at the associate or baccalaureate degree level, excluding specialized and proprietary institutions (n = 2,524). These administrators were instructed to direct the survey to the individual or individuals most familiar with the patterns of undergraduate student assessment within the institution. Completed surveys were received from 1,393 institutions for an overall response rate of 55%. Response rates by institutional type varied from a low of 53% for baccalaureate colleges to a high of 65% for research institutions.

Analyses

We supplemented data collected via our survey with variables for institutional type, control, and enrollment from the 1995 Integrated Postsecondary Education Data System (IPEDS). We used two data reduction approaches to identify patterns of variables in each conceptual domain and to condense the number of variables used in bivariate analyses. Survey items with an interval response scale were factored using an oblique rotation method. Several sections of the survey inventory consisted of dichotomous or categorical variables that did not lend themselves to factor analysis. In most instances, we reduced data by summing scores within sections to create an additive index of practices or policies in existence at institutions.

Descriptive mean scores of the variable indices were used to examine student assessment approaches, organizational and administrative support practices, assessment policies and practices, and uses of student assessment information in all respondent institutions. One-way analysis of variance was used to examine differences in these measures across five types of institutions: associate of arts, baccalaureate, master's, doctoral, and research. For each measure, post hoc multiple comparisons were conducted using the Tukey method to identify significant differences among specific institutional types. The eta-squared statistic was used to measure the proportion of the variability in each measure accounted for by institution type. Impacts of student assessment information were categorical variables. The Pearson chi-square statistic was used to test for significant differences in reported impacts by institutional type. Although not presented in the tables, adjusted residuals were used to identify which cross-tabulation cells departed most from the hypothesis of independence, and standardized residuals were used to determine the contribution of each cell to the chi-square statistic.

Results

Institutional Approaches to Student Assessment

Our first research question concerned the approaches to student assessment adopted by institutions. Eight variables were used to examine institutions' student assessment approaches. Table 1 displays means and standard deviations of assessment approach variables for all respondent institutions and by institutional type.

All institutions. In general, the extent of data collection among all institutions was limited. Institutions collected postcollege assessment data (students' vocational outcomes, further education, and postenrollment satisfaction with institution) more extensively than data concerning current students' cognitive performance or affective development. Institutions reported substantial use of comprehensive tests or instruments as means of collecting assessment data. Despite considerable interest in the literature concerning alternative methods of assessing student performance, student-centered methods such as portfolios and capstone projects were either not used or were used in only some departments. Use of external assessment methods (those involving external constituencies such as employers and alumni) was comparatively greater. Institutions produced and provided reports of student assessment results at several levels of aggregation. It appears these reports were primarily descriptive in nature as institutions reported conducting few studies of the relationship between aspects of students' institutional experiences and student performance.

By institutional type. Statistically significant differences by institutional type were evident on all assessment approach variables except the use of external assessment methods. However, absolute differences among institutional types in the means and standard deviations for each assessment approach measure were generally very small. The eta-squared statistic shows that institutional type accounted for a substantive proportion (greater than 5%) of the variance in only four approach measures—the collection of cognitive, affective and postcollege assessment data and the use of student-centered assessment methods.

TABLE 1

Institutional Approaches to Student Assessment

	Institutional Type							
	All N = 1298	Assoc of Arts N = 539	Bacca-laureate N = 309	Master's N = 311	Doctoral N = 64	Research N = 78	*F*	Eta squared
Content								
Cognitive assessment[1]	1.67 (0.58)	1.57^{b} (0.58)	1.81^{a} (0.58)	1.75^{a} (0.55)	1.69^{ab} (0.55)	1.50^{b} (0.53)	10.881***	0.201
Affective assessment[2]	1.87 (0.54)	1.70^{d} (0.49)	2.06^{ab} (0.58)	1.94^{c} (0.51)	1.86^{abcd} (0.64)	2.01^{abc} (0.53)	26.901***	0.113
Post-college assessment[2]	2.29 (0.60)	2.29^{a} (0.59)	2.32^{a} (0.60)	2.34^{a} (0.59)	2.21^{ab} (0.64)	2.09^{b} (0.60)	3.302*	0.072
Methods								
Number of instruments[3]	9.41 (.3.49)	8.63^{b} (3.33)	9.67^{a} (3.40)	10.41^{a} (3.45)	10.52^{a} (3.49)	9.17^{ab} (3.51)	14.211***	0.045
Student-centered methods[4]	1.37 (0.30)	1.24^{b} (0.27)	1.51^{a} (0.31)	1.46^{a} (0.27)	1.50^{a} (0.28)	1.33^{b} (0.22)	56.501***	0.200
External methods[5]	2.04 (0.57)	2.03 (0.60)	1.98 (0.54)	2.10 (0.53)	2.05 (0.54)	2.05 (0.58)	1.713	0.045
Analyses								
Number of studies[6]	2.22 (2.32)	1.80^{c} (2.12)	2.32^{b} (2.36)	2.44^{b} (2.39)	2.92^{ab} (2.59)	3.33^{a} (2.36)	11.284***	0.048
Number of reports[7]	2.48 (1.46)	2.31^{b} (1.48)	2.18^{b} (1.31)	2.82^{a} (1.42)	2.94^{a} (1.47)	3.13^{a} (1.57)	15.255***	0.045

NOTE: Standard deviations are in parentheses. Post hoc pairwise multiple comparisons were tested using the Tukey method. For each variable, means whose superscripts contain a common letter (e.g., ^{a}) are *not* significantly different ($p > 0.001$); means whose superscripts do not contain a common letter *are* significantly different ($p < 0.001$). Response categories for all additive indices: 1 = yes; 0 = no. Factor scale A: 1 = not collected; 2 = collected for some students; 3 = collected for many students; 4 = collected for all students. Factor scale B: 1 = not used; 2 = used in some units; 3 = used in most units; 4 = used in all units. [1]Four-item factor: scale A. [2]Three-item factor: scale A. [3]Thirty-item additive index. [4]Four-item factor: scale B. [5]Two-item factor: scale B. [6]Nine-item additive index. [7]Five-item additive index.

* $p < 0.05$. *** $p < 0.001$.

Like all types of institutions except research universities, associate of arts colleges collected postcollege data from some to many of their former students. However, they collected significantly fewer affective and cognitive data than baccalaureate and master's institutions and used significantly fewer standardized instruments and student-centered assessment methods than all but research institutions. Associate of arts colleges conducted significantly fewer assessment studies than other institutional types and produced fewer assessment reports than all but baccalaureate colleges.

Baccalaureate and master's institutions shared similar assessment approach profiles. While baccalaureate colleges collected significantly more affective data than master's institutions, both collected significantly more cognitive data than associate of arts and research institutions and significantly more postcollege data than research institutions. They used significantly more standardized instruments than associate of arts colleges and made greater use of student-centered methods than associate of arts and research institutions. Both fell in the mid-range of number of assessment studies conducted. Along with associate of arts institutions, baccalaureate colleges produced significantly fewer reports of assessment data than did other types of institutions.

Doctoral institutions were neither significantly higher nor lower than other institutional types in their collection of cognitive, affective, and postcollege data. Like baccalaureate and master's institutions, they used significantly more standardized instruments than associate of arts colleges and more student-centered assessment methods than associate of arts and research institutions. They conducted significantly more relational studies of assessment data than associate of arts colleges and produced more assessment reports than associate of arts and baccalaureate colleges.

In terms of content and methods, research universities reported the least extensive assessment approaches. They collected significantly fewer postcollege data than all other types of institutions and fewer cognitive data than baccalaureate and master's institutions. They made significantly less use of student-centered assessment methods than all but associate of arts colleges. However, they conducted significantly more studies of relationships between students' institutional experiences and performance than all other institutional types and provided assessment reports at significantly more levels of aggregation than associate of arts and baccalaureate colleges.

Organizational and Administrative Support for Student Assessment

Our second research question addressed the organizational and administrative support patterns that institutions have developed to promote student assessment on campus. Nine variables were used to measure institutional support strategy, leadership and governance for assessment, and evaluation. Table 2 displays means and standard deviations for these variables for all respondent institutions and by institutional type.

All institutions. Institutions reported a moderate degree of mission emphasis on undergraduate education and its assessment. Preparing for accreditation self-study emerged as the most important purpose for engaging in student assessment activities, followed by meeting state reporting requirements and internal improvement. Institutions had introduced two to three administrative and governance activities to promote student assessment. Administrators and faculty were perceived as being somewhat to very supportive of student assessment. Half of respondent institutions had a formal centralized policy stipulating specified student assessment activities of all academic units, almost three-quarters had an institution-wide group responsible for assessment planning and policy development, and half had either formally or informally evaluated their student assessment process.

By institutional type. There were statistically significant differences by institutional type on all dimensions of organizational and administrative support but evaluating the assessment process. With the exception of conducting assessment for state purposes, these differences were small in absolute magnitude and most often reflected differences between research universities and associate of arts, baccalaureate and master's institutions. As shown by the eta-squared statistic, institutional type accounted for more than 5% of the variance in only two organizational and administrative support measures—conducting student assessment for internal purposes and for state purposes.

There were several similarities in the profiles of organizational and administrative support for assessment reported by associate of arts, baccalaureate, and master's institutions. Compared to research institutions, they gave significantly higher importance ratings to internal improvement and preparing for an accreditation self-study as purposes for conducting assessment, reported significantly higher administrative and faculty support for assessment, were more likely to have a formal centralized assessment policy and more likely to use an institution-wide planning group for student assessment.

Differences were also observed among these three institution types. Associate of arts colleges were significantly less likely to have an explicit mission emphasis on undergraduate education and assessment than were baccalaureate and master's institutions. Associate of arts colleges gave significantly greater importance to meeting state requirements as an assessment purpose—rating this as moderately to very

TABLE 2

Organizational and Administrative Support for Student Assessment

	Institutional Type							
	All N = 1312	Assoc of Arts N = 544	Bacca-laureate N = 315	Master's N = 313	Doctoral N = 65	Research N = 79	*F*	Eta squared
Institutional Support Strategy								
Mission emphasis[1]	1.54 (0.85)	1.46^{b} (0.90)	1.64^{a} (0.77)	1.64^{a} (0.84)	1.50^{ab} (0.87)	1.27^{b} (0.73)	5.320***	0.005
Internal purposes[2]	2.49 (0.50)	2.53^{a} (0.51)	2.51^{a} (0.50)	2.48^{a} (0.45)	2.37^{ab} (0.51)	2.28^{b} (0.51)	5.185***	0.138
Accreditation purposes[3]	3.60 (0.65)	3.61^{a} (0.66)	3.63^{a} (0.63)	3.67^{a} (0.55)	3.69^{a} (0.58)	3.14^{b} (0.92)	11.300***	0.007
State purposes[3]	2.88 (1.18)	3.37^{a} (.90)	2.30^{c} (1.21)	2.76^{b} (1.17)	2.60^{bc} (1.26)	2.41^{bc} (1.27)	54.650***	0.064
Leadership and Governance								
Administrative and governance activities[4]	2.34 (1.22)	2.20^{b} (1.18)	2.33^{ab} (1.29)	2.62^{a} (1.21)	2.42^{ab} (1.25)	2.16^{ab} (1.12)	5.212***	0.013
Administrative and faculty support[5]	17.04 (2.74)	17.20^{a} (2.86)	17.09^{a} (2.91)	17.06^{a} (2.37)	16.51^{ab} (2.88)	16.05^{b} (2.27)	3.606**	0.033
Formal centralized policy[6]	0.50 (0.50)	0.54^{a} (0.50)	0.55^{a} (0.50)	0.47^{a} (0.50)	0.37^{ab} (0.49)	0.28^{b} (0.45)	6.951***	0.017
Institution-wide planning group[6]	0.71 (0.46)	0.73^{a} (0.44)	0.73^{a} (0.45)	0.73^{a} (0.45)	0.62^{ab} (0.49)	0.45^{b} (0.50)	6.950***	0.011
Evaluation								
Conducted evaluation[6]	0.49 (0.50)	0.52 (0.50)	0.47 (0.50)	0.49 (0.50)	0.50 (0.50)	0.36 (0.48)	2.039	0.010

NOTE: Standard deviations are in parentheses. Post hoc pairwise multiple comparisons were tested using the Tukey method. For each variable, means whose superscripts contain a common letter (e.g., ^{a}) are *not* significantly different ($p > 0.001$); means whose superscripts do not contain a common letter *are* significantly different ($p < 0.001$). Response categories for all additive indices: 1 = yes; 0 = no. Factor scale A: 1 = no importance; 2 = minor importance; 3 = moderate importance; 4 = very important. [1]Three-item additive index. [2]Four-item factor: scale A. [3]Single item: scale A. [4]Seven-item additive index. [5]Four-item additive index: 1 = very unsupportive; 2 = somewhat unsupportive; 3 = neutral or unknown; 4 = somewhat supportive; 5 = very supportive. [6]Single item: 1 = yes; 0 = no.

** $p < 0.01$. *** $p < 0.001$.

important—than all other types of institutions. They used significantly fewer administrative and governance activities to support assessment than master's institutions.

There were few significant differences between doctoral and other types of institutions on measures of organizational and administrative support for assessment. Doctoral institutions gave a significantly higher importance rating to preparing for an accreditation self-study as a purpose for conducting assessment than did research institutions and attributed significantly less importance than associate of arts institutions to conducting assessment for state purposes.

Research institutions presented the weakest profile of organizational and administrative support for assessment. As discussed above, they scored significantly lower than all but doctoral institutions on most measures considered within this domain.

Assessment Management Policies and Practices

Our third research question addressed the use of assessment management policies and practices to support student assessment. Table 3 displays means and standard deviations of nine dimensions of assessment management policies and practices for all institutions and by institutional type.

All institutions. On the whole, institutions rarely used student performance indicators to reward academic units in the budget process or competitively allocate resources among academic units. Information system capabilities for student assessment, indicated by the computer support index, were limited. Institutions provided fairly broad internal access to student assessment information but somewhat less access through the distribution of assessment reports. Institutions used student performance data in academic planning and review processes and had policies encouraging student involvement in assessment activities in some departments. They made comparatively less use of policies providing professional development on student assessment for faculty, academic administrators, and student affairs personnel and seldom linked faculty evaluation and reward policies to student assessment involvement or results.

By institutional type. With the exception of budget decisions and the distribution of assessment reports, there were statistically significant differences in the extent to which different types of institutions used assessment management policies and practices. Absolute differences among institutional types were slightly larger in this assessment domain than in the approach and support domains. Institutional type differences

TABLE 3

Assessment Management Policies and Practices

	Institutional Type							
	All N = 1293	Assoc of Arts N = 538	Bacca-laureate N = 304	Master's N = 313	Doctoral N = 308	Research N = 78	*F*	Eta squared
Budget decisions[1]	0.08 (0.32)	0.07 (0.29)	0.07 (0.32)	0.10 (0.33)	0.05 (0.22)	0.18 (0.51)	1.440	0.003
Computer support[2]	0.80 (0.90)	0.89^{a} (0.94)	0.66^{b} (0.87)	0.75^{ab} (0.86)	0.69^{ab} (0.79)	0.92^{ab} (0.91)	4.018**	0.003
Access to information[3]	3.48 (1.65)	3.80^{a} (1.53)	3.35^{b} (1.66)	3.34^{b} (1.62)	2.77^{b} (1.85)	2.94^{b} (1.88)	11.776***	0.034
Distribution of reports[4]	2.44 (1.38)	2.54 (1.42)	2.42 (1.24)	2.34 (1.32)	2.31 (1.48)	2.26 (1.66)	1.663	0.016
Academic planning and review[5]	2.78 (0.97)	2.87^{a} (0.99)	2.82^{ab} (0.99)	2.72^{abc} (0.89)	2.48^{bc} (0.95)	2.43^{c} (0.85)	5.089***	0.302
Student involvement[6]	2.64 (0.85)	2.71^{ab} (0.88)	2.80^{a} (0.80)	2.56^{b} (0.83)	2.41^{bc} (0.77)	2.14^{c} (0.79)	11.040**	0.156
Professional development[5]	1.88 (0.79)	2.04^{a} (0.82)	1.79^{b} (0.77)	1.81^{b} (0.76)	1.83^{ab} (0.73)	1.45^{c} (0.60)	11.975***	0.292
Student affairs[7]	1.93 (1.17)	2.21^{a} (1.27)	1.66^{b} (1.07)	1.78^{b} (1.07)	1.80^{ab} (1.08)	1.68^{b} (0.95)	13.505***	0.042
Faculty evaluation[8]	1.25 (0.66)	1.14^{b} (0.62)	1.37^{a} (0.74)	1.31^{a} (0.63)	1.26^{ab} (0.63)	1.22^{ab} (0.59)	6.866***	0.352

NOTE: Standard deviations are in parentheses. Post hoc pairwise multiple comparisons were tested using the Tukey method. For each variable, means whose superscripts contain a common letter (e.g., ^{a}) are *not* significantly different ($p > 0.001$); means whose superscripts do not contain a common letter *are* significantly different ($p < 0.001$). Response categories for all additive indices: 1 = yes; 0 = no. Scale for all factors: 1 = not done at all; 2 = done in a few depts; 3 = done in some depts; 4 = done in many depts; 5 = done in most depts. [1] Two-item additive index. [2] Three-item additive index. [3] Five-item additive index. [4] Six-item additive index. [5] Four-item factor scale. [6] Three-item factor scale. [7] Two-item factor scale. [8] Five-item factor scale.

** $p < 0.01$. *** $p < 0.001$.

were also reflected in the large eta-squared statistics associated with the use of four areas of assessment management policies: academic planning and review, student involvement, professional development for faculty and academic administrators, and faculty evaluation.

Overall, associate of arts institutions made the most extensive use of assessment management policies and practices. They used more sophisticated information systems to manage student assessment data and provided significantly greater internal access to these data than other types of institutions. They made significantly greater use of assessment information in academic planning and review and had significantly more extensive policies encouraging student involvement in assessment than research universities. They had significantly more extensive professional development policies for faculty, academic administrators and student affairs personnel than all types of institutions but doctoral universities. Although it was rare for any type of institution to link student assessment to faculty evaluation and rewards, associate of arts colleges did so least of all institutional types.

There were several similarities in the extent to which baccalaureate, master's, doctoral, and research institutions used these assessment management policies and practices. They did not differ significantly from each other in their computer support for assessment, provision of internal access to assessment information, and the extent of their policies providing professional development on assessment for student affairs personnel or linking student assessment to faculty evaluation. In particular, master's and doctoral institutions fell in the mid-range of responses relative to other types of institutions, often reporting neither significantly more nor less extensive use of assessment policies and practices to support student assessment.

However, a few distinctions among these institutional types were also observed. Baccalaureate colleges reported significantly less computer support for assessment than associate of arts colleges and made greater use of policies promoting student involvement in assessment than master's, doctoral, and research universities. Research universities were significantly less likely to use assessment data in academic planning and review processes than associate of arts and baccalaureate colleges; made significantly less use of policies encouraging student involvement in assessment than all but doctoral institutions; and provided significantly less professional development on student assessment for faculty and academic administrators than all other types of institutions.

Institutional Uses and Impacts of Student Assessment

Assessment information uses. Our fourth research question considered institutional uses and impacts of student assessment information. Two indices were used as measures of the influence of student assessment information on institutional decision making. Academic decisions was a ten-item factor reflecting the influence of student assessment information on decisions such as academic program planning, academic support services, teaching methods, and academic resource allocation. Faculty decisions was a two-item factor measuring the influence of student assessment information on institutions' decisions regarding faculty promotion and tenure and salary increases or other rewards. Specific items and alpha coefficients for each factor appear in Appendix A. Table 4 displays means and standard deviations for these assessment uses for all institutions and by institutional type.

All institutions. As Table 4 shows, student assessment information has had limited influence on institutions' academic decisions and even less on faculty-related decisions. These low scores for assessment information uses reflect the fact that many institutions reported they had not monitored the items comprising these indices (for full results of individual items comprising these indices, see Peterson, Einarson, Augustine, & Vaughan, 1999c). The percentage of institutions reporting they did not know the nature or extent of assessment information influences ranged from 22% for modifications to student academic support services to 70% for faculty salary increases or rewards. When institutions had monitored assessment uses on specific items, they were most likely to report that assessment information had been somewhat influential with respect to decision making.

By institutional type. There were statistically significant differences by institutional type in reported use of assessment information in academic and faculty decisions. For example, post hoc comparisons reveal that master's institutions made significantly greater use of assessment information in academic decisions than research institutions. Baccalaureate colleges reported significantly more influence of assessment data in faculty-related decisions than either associate of arts or research institutions. Further, the eta-squared statistic shows that institutional type accounted for a large proportion of the variance in using assessment information in academic decisions. But statistical significance notwithstanding, absolute differences in the use of assessment information by types of institutions were negligible. All types reported that assessment information had either none or no known influence on academic and faculty decisions.

Assessment information impacts. Four measures of the impact of student assessment on students were used: student satisfaction, retention or graduation rates, grade performance, and achievement on external examinations. Four items served as measures of the impact of student assessment on faculty: campus discussions of undergraduate education, faculty satisfaction, interest in teaching and changes in teaching methods. Four items measured the impact of student assessment information on external indicators of institutional performance: student applications or acceptance rates, allocation of state funding, regional accreditation evaluation, and institutional reputation. For each item, respondents indicated whether they had not monitored the impact of assessment information, had monitored and observed a negative impact, had monitored and observed no impact, or had monitored and observed a positive impact. Table 5 presents the distribution of responses for all institutions and by institutional type. Chi-square was used to test the significance of differences in responses across institutional types.

All institutions. Table 5 clearly shows that the majority of institutions had not monitored assessment-related impacts. Institutions were more likely to have monitored assessment impacts on students and faculty than on external performance indicators. When monitoring had taken place, negative impacts were almost never reported.

Approximately one-fifth of institutions reported positive impacts of assessment information on three indirect indicators of student performance—student satisfaction, retention or graduation rates, and achievement on external examinations. A smaller proportion had documented positive impacts of assessment information on student grade achievement. More than one-third of institutions reported that assessment information had stimulated campus discussions of undergraduate education and encouraged faculty to change their teaching methods, while one-fifth attributed increased faculty interest in teaching to their student assessment efforts. There was a mixed profile of assessment impacts on faculty satisfaction. Whereas one-tenth of institutions reported a positive impact from student assessment information on this impact measure, five percent reported a negative impact; although a small proportion, this was the largest negative response associated with any impact measure.

Conversely, institutions were most likely to report a positive impact from student assessment information on their regional accreditation evaluation—with more than two-fifths of institutions doing so. One-fifth of respondents credited their student assessment efforts with an improvement in institutional reputation. Fully 90% of institutions had either not monitored or did not know the impact of assessment information on student applications or state funding allocations.

By institutional type. There were no statistically significant differences by institution type in the monitoring and nature of assessment impacts on any of the four measures of student performance, faculty interest in teaching, accreditation evaluation and institutional reputation.

TABLE 5

Impacts of Student Assessment Information

	Institutional Type						
	All N = 1302	Assoc of Arts N = 519	Bacca-laureate N = 295	Master's N = 296	Doctoral N = 63	Research N = 69	Chi Square
Student Impacts							12.537
Student satisfaction							
Not monitored	58.1	59.6	53.7	58.3	60.9	62.9	
Monitored/negative impact	1.0	1.1	0.7	1.7	0.0	0.0	
Monitored/no known impact	21.2	20.1	26.5	17.7	21.9	21.4	
Monitored/positive impact	19.6	19.2	19.1	22.3	17.2	15.7	
Retention/graduation rates							8.127
Not monitored	49.9	49.0	46.8	51.8	57.1	55.1	
Monitored/negative impact	0.8	0.9	1.0	0.7	0.0	0.0	
Monitored/no known impact	28.6	27.1	31.6	28.6	28.6	27.5	
Monitored/positive impact	20.7	23.0	20.5	18.9	14.3	17.4	
Grade performance							17.992
Not monitored	58.9	54.0	59.1	64.6	65.6	64.3	
Monitored/negative impact	0.6	0.8	1.0	0.3	0.0	0.0	
Monitored/no known impact	27.6	28.6	29.2	25.2	25.0	25.7	
Monitored/positive impact	12.8	16.6	10.6	9.9	9.4	10.0	
External achievement							9.783
Not monitored	60.8	59.7	59.1	62.5	61.5	68.1	
Monitored/negative impact	0.3	0.2	0.7	0.3	0.0	0.0	
Monitored/no known impact	19.8	19.3	22.1	18.3	16.9	23.2	
Monitored/positive impact	19.1	20.8	18.2	18.9	21.5	8.7	

TABLE 5 (*Continued*)

	Institutional Type						
	All N = 1302	Assoc of Arts N = 519	Bacca-laureate N = 295	Master's N = 296	Doctoral N = 63	Research N = 69	Chi Square
Faculty Impacts							
Campus discussions							29.833**
Not monitored	51.5	57.5	42.1	48.0	54.7	59.2	
Monitored/negative impact	1.0	1.1	1.0	1.3	0.0	0.0	
Monitored/no known impact	13.9	13.6	14.6	12.6	18.8	14.1	
Monitored/positive impact	33.6	27.8	42.4	38.1	26.6	26.8	
Faculty satisfaction							31.330**
Not monitored	67.9	68.2	58.9	70.7	75.0	85.5	
Monitored/negative impact	5.3	4.2	7.7	6.1	1.6	4.3	
Monitored/no known impact	16.7	16.0	20.4	16.2	15.6	8.7	
Monitored/positive impact	10.1	11.6	13.0	7.1	7.8	1.4	
Interest in teaching							20.520
Not monitored	65.3	65.0	60.7	66.4	71.4	77.1	
Monitored/negative impact	1.3	1.9	1.0	1.0	0.0	1.4	
Monitored/no known impact	13.8	15.3	17.7	9.6	11.1	5.7	
Monitored/positive impact	19.6	17.8	20.7	22.9	17.5	15.7	
Teaching methods							22.206*
Not monitored	46.2	47.4	39.8	47.5	43.8	61.4	
Monitored/negative impact	0.2	0.4	0.0	0.0	0.0	0.0	
Monitored/no known impact	15.7	15.8	21.1	11.9	15.6	8.6	
Monitored/positive impact	37.9	36.4	39.1	40.6	40.6	30.0	

TABLE 5 (*Continued*)

	Institutional Type						
	All N = 1302	Assoc of Arts N = 519	Bacca-laureate N = 295	Master's N = 296	Doctoral N = 63	Research N = 69	Chi Square
External Impacts							
Student applications							21.674*
Not monitored	78.8	82.6	72.2	78.6	81.3	76.4	
Monitored/negative impact	0.8	1.0	0.3	1.0	0.0	1.4	
Monitored/no known impact	13.4	9.9	20.1	12.0	12.5	16.7	
Monitored/positive impact	7.1	6.5	7.4	8.4	6.3	5.6	
State funding							30.188**
Not monitored	80.5	77.3	89.5	80.8	75.4	72.1	
Monitored/negative impact	1.1	1.3	0.4	1.4	1.6	0.0	
Monitored/no known impact	10.6	10.6	6.9	11.8	13.1	17.6	
Monitored/positive impact	7.8	10.8	3.2	5.9	9.8	10.3	
Accreditation evaluation							13.091
Not monitored	42.5	45.1	41.4	38.9	37.1	48.5	
Monitored/negative impact	2.8	2.5	2.4	3.4	3.2	4.4	
Monitored/no known impact	12.5	12.5	14.6	10.2	9.7	16.2	
Monitored/positive impact	42.1	39.9	41.7	47.4	50.0	30.9	
Institutional reputation							14.109
Not monitored	63.7	65.6	59.5	62.0	71.0	69.0	
Monitored/negative impact	0.6	0.4	0.7	1.0	0.0	1.4	
Monitored/no known impact	13.7	11.7	16.4	13.3	16.1	16.9	
Monitored/positive impact	21.9	22.4	23.4	23.7	12.9	12.7	

NOTE: Due to rounding column percentages for each impact item may not total 100%. Differences by institutional type were estimated using chi-square test of independence.
$*p < 0.05$. $**p < 0.01$.

Standardized and adjusted residuals were examined to identify cells making the greatest contribution to significant institutional type differences for the remaining faculty and external impact measures.

Institution types differed significantly on three of the four faculty impact items: campus discussions of undergraduate education, faculty satisfaction, and changes in teaching methods. The significant chi-square statistics associated with these three items were chiefly the product of the greater tendency of baccalaureate institutions, relative to other types of institutions, to have monitored and documented positive assessment impacts. Associate of arts, master's, and doctoral institutions were in the mid-range of responses, while research institutions reported monitoring and documenting positive impacts on these faculty-related measures less often than would be expected if monitoring of these impacts were independent of institution type.

Institution types differed significantly on two measures of external impacts from student assessment information: student application or acceptance rates and state funding allocations. Significant chi-square statistics were primarily due to the difference in the responses of associate of arts and baccalaureate colleges on these impact measures. Baccalaureate institutions were more likely than associate of arts institutions to have monitored the impact of assessment information on student application rates, and to have documented no associated impact. However, associate of arts institutions were more likely than baccalaureate institutions to have monitored the impact of assessment on their state funding allocations and to have documented a related positive impact. The other three types of institutions generally fell between these two extremes of responses.

Discussion

Several limitations of the study must be noted. Like much survey research, our data consists of institutions' self-reports of assessment practices, policies, uses and impacts. Our survey was initially directed to the chief academic officer within each institution. Particularly in larger institutions, it would be difficult for any one individual to be knowledgeable about the wide range of institutional policies and practices tapped by survey questions. However, these administrators were instructed to redirect the survey for completion by other individuals within the institution as appropriate. In our request for the names of individuals to serve as future points of contact for survey results, we most often received the names of individuals other than the chief academic officer—usually an institutional research or student assessment officer. This suggests that chief academic officers did indeed redirect the survey for completion as needed and enhances the credibility of the responses we received. An additional limitation of our research was its focus on an institution-wide perspective. This did not permit the examination of variations in assessment practices and policies within individual programs or departments or at the classroom level.

Despite these limitations, this study contributes a current and comprehensive national portrait of the nature and extent of undergraduate student assessment activities, institutional support practices, and uses and impacts within and across specific types of postsecondary institutions and allows a comparison of recommended to actual practices for student assessment among postsecondary institutions. Finally, these findings suggest future research directions.

Institutional Approaches to Student Assessment

Our first research question addressed the nature of student assessment approaches institutions have adopted. Scholars contend that the comprehensiveness of an institution's assessment approach may be positively related to the degree of internal support for assessment and the uses and impacts of assessment information achieved.

Institutions have adopted assessment approaches that are less extensive than those recommended by assessment scholars. Our findings confirm prior research regarding the moderate scope of assessment approaches undertaken by institutions. On the whole, institutions emphasize the collection of easily quantifiable postcollege measures, such as employment outcomes and further education, over more complex measures, such as higher-order cognitive skills and affective development. They make greater use of traditional assessment methods, such as standardized instruments, than less traditional methods, such as portfolios or capstone courses. Although they provide descriptive profiles of student assessment results, they seldom study the relationship between students' institutional experiences and student performance.

We found statistically significant differences in assessment approach measures by institutional type, although absolute differences were of little practical significance for several measures. Institution-type differences in student assessment approaches appear to be reflective of related variations in institutional mission (Austin, 1994; Clark, 1987). Despite their institutional focus on undergraduate education, associate of arts institutions have undertaken the least comprehensive assessment approaches. The highly diverse and mobile student population served by these open admission institutions may render the assessment and analysis of student performance more difficult than in other types of institutions. In contrast, baccalaureate institutions have the most comprehensive data collection efforts. This may be related to the student-development focus of these institutions. Their comparatively low emphasis on data analysis and reporting may be indicative of lower institutional research capacity relative to graduate institutions. It also suggests that student assessment may occur more often at the classroom or department level than as an institution-wide activity. Master's and doctoral institutions both scored in the middle range for most assessment approach variables. This may be due to these institutions' competing responsibilities toward undergraduate and graduate education. Finally, research institutions' low scores for data collection and assessment methods are likely associated with their greater emphasis on graduate rather than undergraduate education. Their high scores for assessment studies and reports reflect a well-developed capacity for educational research.

Organizational and Administrative Support for Student Assessment

Our second research question examined the organizational and administrative support patterns used by institutions to promote student assessment. Scholars and practitioners of student assessment assert that conducting assessment for internal purposes, strong leadership support, a combination of centralized and decentralized governance structures, and ongoing evaluation of assessment efforts are important means by which institutions can effectively support and benefit from student assessment.

A mixed profile of support practices was observed across all institutions. As has been noted in previous studies (El-Khawas, 1992, 1995), regional accreditation requirements appear to be responsible for stimulating many institutions' assessment efforts; whereas meeting state requirements and informing internal improvement are comparatively less important purposes underlying institutions' decisions to engage in assessment. There is limited evidence of explicit leadership support for assessment through mission emphasis or administrative/governance activities. Many institutions have established formal centralized policies concerning student assessment, and most utilize some degree of internal participation in assessment planning. Only half of institutions have evaluated their assessment process.

Significant differences by institutional type were evident in patterns of organizational and administrative support practices for student assessment. These were largest for the importance accorded to state-related purposes for conducting assessment. Differences reflect the governance styles associated with specific types of institutions (Austin &

Gamson, 1983; Birnbaum, 1988, Clark, 1987). For associate of arts and baccalaureate colleges, institutional control also plays a role in shaping support patterns.[1] Associate of arts institutions have a comparatively strong profile of organizational and administrative support practices—an emphasis that is congruent with their more bureaucratic or administratively driven governance style relative to other types of institutions (Cohen & Brawer, 1996). These institutions view assessment as serving both internal and external purposes. The high importance attributed to meeting state requirements is attributable to the large proportion of these institutions under public control. Baccalaureate institutions are more often privately controlled institutions and thus place the least emphasis on fulfilling state-related purposes for assessment. Together with master's institutions, they fall in the middle range of administrative control and faculty autonomy. In contrast, research universities and, to a lesser extent, doctoral universities operate as professional bureaucracies—institutions in which management functions and external accountability are the responsibility of administrators, while teaching and scholarship are the responsibility of an autonomous faculty. The governance style and emphasis on graduate education characteristic of these institutions likely contributes to their comparatively low scores on organizational and administrative support practices for undergraduate student assessment.

Assessment Management Policies and Practices

Our third research question considered the extent to which institutions have intentionally used specific policies and practices to support student assessment. Scholars have depicted assessment management policies and practices as powerful levers for integrating student assessment with other institutional functions and processes, thereby fostering internal support for and institutional impacts from assessment.

Institutions have provided moderate internal access to and distribution of student assessment information. There appear to be some efforts to integrate student assessment processes with academic planning and review processes. Many institutions have also adopted policies to encourage student involvement in assessment activities. However, there is less evidence of providing professional development on student assessment to faculty, academic administrators, and student affairs personnel, and almost no efforts to link student assessment with faculty evaluation and reward policies.

Observed differences in the use of assessment management policies and practices among institutional types are consistent with institutional missions and governance styles. Of all types of institutions, associate of arts institutions are most likely to link student assessment to academic planning processes and to provide professional development on assessment to faculty and administrators. This administrative focus on assessment support is congruent with the comparatively strong administrative control associated with these institutions. Baccalaureate institutions, on the other hand, are more likely to have policies encouraging faculty and student involvement in assessment. This profile is reflective of their strong student orientation and suggests that student assessment in these institutions is more often a faculty than an administrative or student affairs responsibility. The relatively moderate use of these policies and practices by master's and doctoral institutions seems characteristic of their dual focus on graduate and undergraduate education. Finally, in keeping with their emphasis on graduate education and high degree of faculty autonomy, research institutions present the least extensive profile of assessment management policies and practices in support of undergraduate student assessment.

Uses and Impacts of Student Assessment

Our fourth research question examined the institutional uses and impacts of student assessment. Assessment scholars clearly contend that the assessment of student performance should not be an end in itself but a means to educational and institutional improvement (AAHE, 1992; Banta & Associates, 1993; Ewell, 1988b).

As has been reported elsewhere in the literature (cf. Banta et al., 1993; Banta & Associates, 1996), most colleges and universities have not monitored the influence of student assessment information on institutional decision making or its impact on student performance, faculty behavior, and external performance indicators. In comparative terms, baccalaureate institutions are most likely of all institutional types to monitor assessment uses and impacts, but in absolute terms, even their rate of reported monitoring of impacts is low. To date, the majority of postsecondary institutions have little documented evidence of whether and to what degree their assessment efforts have influenced decision making or produced discernible impacts on students, faculty, or external constituencies.

Institutions may be doing little monitoring of assessment uses and impacts for a variety of reasons—among them the brief amount of time in which some institutions have been engaged in assessment efforts, the difficulty of disentangling assessment influences and impacts from those of other sources, and the dearth of external requirements to provide evidence of having used assessment information or improved as a consequence of assessment efforts. Given the extensive claims made for the value of student assessment and the substantial human and financial resources invested in assessment activities, institutions need to give greater priority to examining and evaluating the uses and impacts of student assessment.

Future Research

This study provides a systematic comparative examination of institutional approaches to, support practices for, and uses and impacts of undergraduate student assessment. Despite more than a decade of assessment scholarship and activity, results show that most institutions have adopted limited student assessment approaches, developed selected institutional mechanisms to support and promote assessment, and have not monitored the uses and impacts of their assessment efforts. Results also show that patterns of assessment approaches, support practices, and uses and impacts vary among types of postsecondary institutions.

The present study has not examined the relationship of assessment approaches, organizational and administrative support practices, and assessment management policies and practices to institutional uses and impacts of assessment information. Certainly, this is an important research goal. Quantitative multivariate analyses of our survey data have been used to test the relationships among these domains in the conceptual framework. Results of these analyses are available elsewhere (see Peterson & Augustine, 2000a, 2000b).

The present study also supports the importance of conducting analyses within specific types of institutions. Such examinations of institutional assessment approaches, organizational and administrative support patterns, assessment management policies and practices, uses and impacts of student assessment, and relationships among these domains have been conducted. Reports for associate of arts (Peterson, Augustine, Einarson & Vaughan, 1999a), baccalaureate (Peterson, Augustine, Einarson & Vaughan, 1999b), master's (Peterson, Einarson, Augustine, & Vaughan, 1999a) and research and doctoral (Peterson, Einarson, Augustine, & Vaughan, 1999b) institutions are available from the National Center for Postsecondary Improvement.

A qualitative research approach would illuminate institutional processes involved in promoting and supporting student assessment. Conducting case studies of institutions with comprehensive student assessment processes that have produced institutional impacts may help identify institutional best practices in student assessment. A final phase of our research takes this approach and will provide further guidance to institutions in selecting assessment approaches, organizational and administrative practices, and assessment management policies that can promote the use of student assessment information and contribute to enhanced performance.

Finally, the literature on student assessment has highlighted the critical role of institutional culture and climate for student assessment in promoting its use to improve institutional performance (Banta & Associates, 1993; Banta et al., 1996; Jones & Ewell, 1993). Research on this domain of institutional dynamics is extremely important. Although not addressed as part of this national survey, developing an instrument for assessing culture and climate is an important facet of our case studies in the final phase of this project.

APPENDIX A
Variable Definitions

Variable	Definition
Institutional Characteristics	
Institutional type	Five-item nominal variable: associate of arts, baccalaureate, master's, doctoral, and research institutions.
Institutional Approach to Student Assessment	
Cognitive assessment	Four-item factor scale reflecting the extent to which an institution collects data on current students' cognitive performance: competence in major field; general education competencies; higher-order cognitive skills; vocational or professional skills (1 = not collected; 2 = collected for some students; 3 = collected for many students; 4 = collected for all students); Cronbach alpha = 0.71.
Affective assessment	Three-item factor scale reflecting the extent to which an institution collects data on current students' affective development and satisfaction: experiences and involvement with institution; satisfaction with institution; personal growth and affective development (1 = not collected; 2 = collected for some students; 3 = collected for many students; 4 = collected for all students); Cronbach alpha = 0.68.
Postcollege assessment	Three-item factor scale reflecting the extent to which an institution collects data from former students: vocational or professional outcomes; further education; satisfaction and experiences with institution (1 = not collected; 2 = collected for some students; 3 = collected for many students; 4 = collected for all students); Cronbach alpha = 0.83.
Number of instruments	Thirty-item additive index of comprehensive examinations, inventories or standardized instruments (institutionally developed, state provided, and commercially available) used by institution to collect ten types of assessment information: student plans or expectations; basic college-readiness skills; higher-order cognitive skills; general education competencies; competence in major; vocational or professional skills; personal growth and affective development; experiences or involvement with institution; satisfaction with institution (1 = instrument used; 0 = instrument not used).
Student-centered methods	Four-item factor scale reflecting the extent to which an institution uses integrative or nontraditional assessment methods: performance in capstone courses; portfolios or comprehensive projects; observations of student performance; individual interviews or focus groups (1 = not used; 2 = used in some units; 3 = used in most units; 4 = used in all units); Cronbach alpha = 0.61.
External methods	Two-item factor scale reflecting the extent to which an institution uses assessment methods that collect data from external constituencies: employer interviews or focus groups; alumni interviews or focus groups (1 = not used; 2 = used in some units; 3 = used in most units; 4 = used in all units); Cronbach alpha = 0.63.
Number of studies	Nine-item additive index of the number of studies an institution conducts on the relationship between student performance and institutional experiences: course-taking patterns; exposure to different teaching methods; student-faculty interaction; extra-curricular activities; residence arrangements; financial aid and/or employment; admission standards or policies; academic advising patterns; classroom, library and/or computing resources (1 = conduct study; 0 = do not conduct study).
Number of reports	Five-item additive index of the levels of aggregation at which student assessment data are provided as reports: institution-wide; schools or colleges; academic programs or departments; special populations or subgroups of students; by course or groups of courses (1 = report provided; 0 = report not provided).

APPENDIX A (*Continued*)

Variable	Definition
Organizational and Administrative Support	
Mission emphasis	Three-item additive index reflecting an institution's mission statement emphasis on undergraduate education and its assessment: emphasizes excellence in undergraduate education; identifies educational outcomes intended for students; refers to student assessment as important activity (1 = yes; 0 = no).
Internal purposes	Four-item factor scale reflecting the importance of internal institutional purposes for undertaking student assessment: guiding undergraduate academic program improvement; improving achievement of undergraduate students; improving faculty instructional performance; guiding resource allocation decisions (1 = no importance; 2 = minor importance; 3 = moderate importance; 4 = very important); Cronbach alpha = 0.79.
Accreditation purposes	Single item: importance of preparing for institutional accreditation self-study as a purpose for undertaking student assessment (1 = no importance; 2 = minor importance; 3 = moderate importance; 4 = very important).
State purposes	Single item: importance of meeting state reporting requirements as a purpose for undertaking student assessment (1 = no importance; 2 = minor importance; 3 = moderate importance; 4 = very important).
Administrative and governance activities	Seven-item additive index reflecting the number of administrative or governance activities used by an institution to promote student assessment: annual institution-wide assessment forums or seminars; rewards or incentives for administrators promoting use of assessment in unit; incentives for academic units to use assessment information; assessment workshops for administrators; board of trustees committee addresses assessment; faculty governance committee addresses assessment; student representation on assessment committees (1 = yes; 0 = no).
Administrative and faculty support	Four-item additive index reflecting the degree to which chief executive officer, academic and student affairs administrators, and faculty support student assessment (1 = very unsupportive; 2 = somewhat unsupportive; 3 = neutral or unknown; 4 = somewhat supportive; 5 = very supportive).
Formal centralized policy	Single item: institution has formal institutional plan or policy requiring specified student assessment activities of all academic units or programs (1 = yes; 0 = no).
Institution-wide planning group	Single item: institution has institution-wide group for student assessment planning and policy setting (1 = yes; 0 = no).
Conducted evaluation	Single item: institution has formally or informally evaluated its student assessment process (1 = yes; 0 = no).
Assessment Management Policies and Practices	
Budget decisions	Two-item additive index reflecting formal use of assessment information in the budget process: to competitively allocate resources among academic units; to reward academic units for improvement (1 = yes; 0 = no).
Computer support	Three-item additive index reflecting institutional capacity to collect and manage student assessment information: computerized student information system includes student performance indicators; student information system tracks individual students; student assessment database integrated with other institutional databases (1 = yes; 0 = no).

APPENDIX A (*Continued*)

Variable	Definition
Access to information	Five-item additive index reflecting internal accessibility of assessment information on individual students by: institutional research or assessment professionals; senior academic administrators; department chairs or academic program administrators; student affairs professionals; faculty advisors (1 = yes; 0 = no).
Distribution of reports	Six-item additive index reflecting the number of constituent groups to whom student assessment reports are regularly distributed: students; faculty; academic administrators; student affairs professionals; employers; general public (1 = yes; 0 = no).
Academic planning and review	Four-item factor scale reflecting the incorporation of student assessment data into academic planning and review processes for: academic departments or undergraduate programs; general education or core curriculum; courses; student academic support services (1 = not done at all; 2 = done in a few departments; 3 = done in some departments; 4 = done in many departments; 5 = done in most departments); Cronbach alpha = 0.84.
Student involvement	Three-item factor scale reflecting the extent to which an institution has policies or practices to promote student involvement in assessment activities: inform students about assessment purposes and uses; require students to participate in assessment activities; provide students with individual feedback on assessment results (1 = not done at all; 2 = done in a few departments; 3 = done in some departments; 4 = done in many departments; 5 = done in most departments); Cronbach alpha = 0.69.
Professional development	Four-item factor scale reflecting existence of professional development policies or practices on student assessment for faculty and academic administrators: provide funds for faculty to attend or present at assessment conferences; offer student assessment workshops or consultation for faculty; provide assistance (e.g., paid leaves, stipends, course reduction) to improve faculty use of student assessment; provide student assessment workshops for academic administrators (1 = not done at all; 2 = done in a few departments; 3 = done in some departments; 4 = done in many departments; 5 = done in most departments); Cronbach alpha = 0.77.
Student affairs	Two-item factor scale reflecting existence of professional development policies or practices on student assessment for student affairs personnel: require assessment training for student affairs staff; provide student assessment workshops for student affairs administrators (1 = not done at all; 2 = done in a few departments; 3 = done in some departments; 4 = done in many departments; 5 = done in most departments); Cronbach alpha = 0.84.
Faculty evaluation	Five-item factor scale reflecting existence of faculty evaluation and reward policies and practices related to student assessment: promotion evaluation considers evidence of student performance; salary evaluation considers evidence of student performance; promotion, tenure or salary reviews consider faculty participation in student assessment; promotion, tenure or salary reviews consider scholarship on assessment; public recognition or awards for faculty use of student assessment (1 = not done at all; 2 = done in a few departments; 3 = done in some departments; 4 = done in many departments; 5 = done in most departments); Cronbach alpha = 0.77.

APPENDIX A (*Continued*)

Variable	Definition
Institutional Uses and Impacts of Student Assessment	
Academic decisions	Ten-item factor scale reflecting the use of student assessment information in academic decisions: modifying instructional or teaching methods; designing or reorganizing academic programs or majors; revising or modifying general education curriculum; creating or modifying out-of-class learning experiences; revising undergraduate academic mission; modifying student academic support services; modifying student assessment plans or processes; designing or reorganizing student affairs units; allocating resources to academic units; creating or modifying distance learning initiatives (1 = no action or influence unknown; 2 = action taken, data not influential; 3 = action taken, data somewhat influential; 4 = action taken, data very influential); Cronbach alpha = 0.83.
Faculty decisions	Two-item factor scale reflecting the use of student assessment information in faculty decisions: deciding faculty promotion and tenure; deciding faculty salary increases or rewards (1 = no action or influence unknown; 2 = action taken, data not influential; 3 = action taken, data somewhat influential; 4 = action taken, data very influential); Cronbach alpha = 0.79.
Faculty impacts	Four single items reflecting student assessment impacts on faculty: affected campus discussions of undergraduate education; contributed to faculty satisfaction; contributed to faculty interest in teaching; led to changes in teaching methods used (1 = not monitored, do not know; 2 = monitored, negative impact; 3 = monitored, no known impact; 4 = monitored, positive impact).
Student impacts	Four single items reflecting student assessment impacts on students: contributed to student satisfaction; affected student retention or graduation rates; affected student grade performance; affected student achievement on external examinations (1 = not monitored, do not know; 2 = monitored, negative impact; 3 = monitored, no known impact; 4 = monitored, positive impact).
External impacts	Four single items reflecting student assessment impacts on external indicators of institutional performance: student applications or acceptance rates; allocation of state funding; regional accreditation evaluations; and institutional reputation or image; (1 = not monitored, do not know; 2 = monitored, negative impact; 3 = monitored, no known impact; 4 = monitored, positive impact).

Notes

[1]Separate analyses, not reported here, revealed differences by institutional control within institutional type on study measures. Control-related differences were most salient to associate of arts colleges (93% in our study were public) and baccalaureate colleges (88% in our study were private).

References

Alexander, J. M., & Stark, J. S. (1986). *Focusing on student academic outcomes: A working paper* (Technical Report No. 86-A-002.1). Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

American Association for Higher Education. (1992). *Principles of good practice for assessing student learning.* Washington, DC: American Association for Higher Education.

Aper, J. P., Cuver, S. M., & Hinkle, D. E. (1990). Coming to terms with the accountability versus improvement debate in assessment. *Higher Education, 20,* 471–483.

Astin, A. W. (1987). Assessment, value-added, and educational excellence. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain* (New Directions for Institutional Research No. 84, pp. 47-63). San Francisco: Jossey-Bass.

Astin, A. W. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education.* New York: American Council on Education/Macmillan.

Austin, A. E. (1994). Understanding and assessing faculty cultures and climates. In M. K. Kinnick (Ed.), *Providing useful information for deans and department chairs* (New Directions for Institutional Research No. 84, pp. 47-63). San Francisco: Jossey-Bass.

Austin, A. E., & Gamson, Z. F. (1983). *Academic workplace: New demands, heightened tensions* (ASHE-ERIC Higher Education Report No. 10). Washington, DC: Association for the Study of Higher Education.

Banta, T. W. (1985). Using outcomes information at the University of Tennessee, Knoxville. In P. T. Ewell (Ed.). *Assessing educational outcomes* (New Directions for Institutional Research, No. 47, pp. 19–32). San Francisco: Jossey-Bass.

Banta, T. W., & Associates. (1993). *Making a difference: Outcomes of a decade of assessment in higher education.* San Francisco: Jossey-Bass.

Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (1996). *Assessment in practice: Putting principles to work on college campuses.* San Francisco: Jossey-Bass.

Banta, T. W., & Moffett, M. S. (1987). Performance funding in Tennessee: Stimulus for program improvement. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain* (New Directions for Higher Education No. 59, pp. 35-43). San Francisco: Jossey-Bass.

Barak, R. J., & Sweeney, J. D. (1995). Academic program review in planning, budgeting, and assessment. In R. J. Barak & L. A. Mets (Eds.), *Using academic program review* (New Directions for Institutional Research No. 86, pp. 3–17). San Francisco: Jossey-Bass.

Birnbaum, R. (1988). *How colleges work: The cybernetics of academic organization and leadership.* San Francisco: Jossey-Bass.

Bowen, H. R. (1977). *Investment in learning.* San Francisco: Jossey-Bass.

Bowyer, K. A. (1996). Efforts to continually improve a nursing program. In T. W. Banta, J. P. Lund, K. E. Black, & F. W. Oblander (Eds.), *Assessment in practice: Putting principles to work on college campuses* (pp. 128-129). San Francisco: Jossey-Bass.

Braskamp, L. A. (1991). Purposes, issues, and principles of assessment. *NCA Quarterly, 66,* 417-429.

California State University Institute for Teaching and Learning. (1993). *Academic challenges: Student outcomes assessment.* Long Beach: California State University, Institute for Teaching and Learning.

Chaffe-Stengel, P. (1992). Integrated program review: Lessons from assessment research at California State University, Fresno, *Student outcomes assessment: What makes it work? Assessment practices and experiences in the California State University.* Long Beach: California State University Institute for Teaching and Learning.

Clark, B. (1987). *The academic life: Small worlds, different worlds.* Princeton, NJ: Carnegie Endowment for the Advancement of Teaching.

Cohen, A. M., & Brawer, F. B. (1996). *The American community college* (3rd ed.). San Francisco: Jossey-Bass.

Cowart, S. C. (1990). *A survey on using student outcomes measures to assess institutional effectiveness.* Iowa City, IA: American College Testing Program.

Dennison, G. M., & Bunda, M. A. (1989). Assessment and academic judgments in higher education. In P. J. Gray (Ed.), *Achieving assessment goals using evaluation techniques* (New Directions for Higher Education No. 67, pp. 51–70). San Francisco: Jossey-Bass.

Duvall, B. (1994). Obtaining student cooperation. In T. H. Bers & M. L. Mittler (Eds.), *Assessment and testing: Myths and realities* (New Directions for Community Colleges No. 88, pp. 47-52). San Francisco: Jossey-Bass.

Eisenman, C. D. (1991). Faculty participation in assessment programs. *NCA Quarterly, 66,* 458-464.

El-Khawas, E. (1988). *Campus trends 1988.* Higher Education Panel Report No. 77. Washington, DC: American Council on Education.

El-Khawas, E. (1989). How are assessment results being used? *Assessment Update, 1,* (4), 1–2.

El-Khawas, E. (1990). *Campus trends 1990.* Higher Education Panel Report No. 80. Washington, DC: American Council on Education.

El-Khawas, E. (1992). *Campus trends 1992.* Higher Education Panel Report No. 82. Washington, DC: American Council on Education.

El-Khawas, E. (1995). *Campus trends 1995.* Higher Education Panel Report No. 85. Washington, DC: American Council on Education.

Erwin, T. D. (1991). New opportunities: How student affairs can contribute to outcomes assessment. In U. Delworth, G. R. Hanson, & Associates (Eds.), *Student services: A handbook for the profession* (2nd ed., pp. 584–603). San Francisco: Jossey-Bass.

Ewell, P. T. (1984). *The self-regarding institution: Information for excellence.* Boulder, CO: National Center for Higher Education Management Systems.

Ewell, P. T. (1987a). *Assessment, accountability, and improvement: Managing the contradiction.* Boulder, CO: National Center for Higher Education Management Systems.

Ewell, P. T. (1987b). Assessment: Where are we? The implications of new state mandates. *Change, 19*(1), 23–28.

Ewell, P. T. (1987c). Establishing a campus-based assessment program. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain* (New Directions for Higher Education No. 59, pp. 9–24). San Francisco: Jossey-Bass.

Ewell, P. T. (1988a). Implementing assessment: Some organizational issues. In T. W. Banta (Ed.), *Implementing outcomes assessment: Promise and perils* (New Directions for Institutional Research No. 59, pp. 15-28). San Francisco: Jossey-Bass.

Ewell, P. T. (1988b). Outcomes, assessment, and academic improvement: In search of usable knowledge. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 4, pp. 53–108). New York: Agathon Press.

Ewell, P. T. (1991). Assessment and public accountability: Back to the future. *Change, 23*(6), 12–17.

Ewell, P. T. (1997). Strengthening assessment for academic quality improvement. In M. W. Peterson, D. D. Dill, L. A. Mets, & Associates (Eds.), *Planning and management for a changing environment: A handbook on redesigning postsecondary institutions* (pp. 360–381). San Francisco: Jossey-Bass.

Fong, B. (1988). Assessing the departmental major. In J. H. McMillan (Ed.), *Assessing students' learning* (New Directions for Teaching and Learning No. 34, pp. 71–83). San Francisco: Jossey-Bass.

Friedlander, J., Murrell, P. H., & MacDougall, P. R. (1993). The community college student experiences questionnaire. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 196–210). San Francisco: Jossey-Bass.

Gill, W. E. (1993, June 12, 1993). *Conversations about accreditation: Middle States Association of Colleges and Schools focusing on outcomes assessment in the accreditation process.* Paper presented at the Double Feature Conference on Assessment and Continuous Quality Improvement of the American Association for Higher Education, Chicago, IL.

Gentemann, K. M., Fletcher, J. J., & Potter, D. L. (1994). Refocusing the academic program review on student learning. In M. K. Kinnick (Ed.), *Providing useful information for deans and department chairs* (New Directions for Institutional Research No. 84, pp. 31–46). San Francisco: Jossey-Bass.

Gray, P. J., & Banta, T. W. (Eds.). (1997). *The campus-level impact of assessment: Progress, problems and possibilities* (New Directions for Higher Education No. 100). San Francisco: Jossey-Bass.

Halpern, D. F. (Ed.). (1987). *Student outcomes assessment: What institutions stand to gain* (New Directions for Higher Education No. 59). San Francisco: Jossey-Bass.

Hutchings, P., & Marchese, T. (1990). Watching assessment: Questions, stories, prospects. *Change, 22*(5), 12–38.

Hyman, R. E., Beeler, K. J., & Benedict, L. G. (1994). Outcomes assessment and student affairs: New roles and expectations. *NASPA Journal, 32*(1), 20–30.

Jacobi, M., Astin, A., & Ayala, F. (1987). *College student outcomes assessment: A talent development perspective* (ASHE-ERIC Higher Education Report No. 7). Washington, DC: Association for the Study of Higher Education.

Johnson, R., McCormick, R. D., Prus, J., S., & Rogers, J., S. (1993). Assessment options for the college major. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 151–167). San Francisco: Jossey-Bass.

Johnson, R., Prus, J., Andersen, C. J., & El-Khawas, E. (1991). *Assessing assessment: An in-depth status report on the higher education assessment movement in 1990.* Higher Education Panel Report No. 79. Washington, DC: American Council on Education.

Jones, D. P., & Ewell, P. (1993). *The effect of state policy on undergraduate education: State policy and college learning.* Boulder, CO: Education Commission of the States.

Katz, A. M. (1993). Helping a campus in transition. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 54–65). San Francisco: Jossey-Bass.

Krueger, D. W., & Heisserer, M. L. (1987). Assessment and involvement: Investments to enhance learning. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain* (New Directions for Higher Education No. 59, pp. 45-56). San Francisco: Jossey-Bass.

Lenning, O. T., Lee, Y. S., Micek, S. S., & Service, A. L. (1977). *A structure for the outcomes of postsecondary education.* Boulder, CO: National Center for Higher Education Management Systems.

Loacker, G., & Mentkowski, M. (1993). Creating a culture where assessment improves learning. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 5-24). San Francisco: Jossey-Bass.

Magruder, W. J., & Young, C. C. (1993). Value-added talent development in general education. In T. W. Banta, J. P. Lund, K. E. Black, & F. W. Oblander (Eds.), *Assessment in practice: Putting principles to work on college campuses* (pp. 169–171). San Francisco: Jossey-Bass.

McClain, C. J., Krueger, D. W., & Taylor, T. (1986). Northeast Missouri State University value-added assessment program: A model for educational accountability. *International Journal of Institutional Management in Higher Education, 10*(3), 252–261).

Mentkowski, M. (1991). Creating a context where institutional assessment yields educational improvement. *Journal of General Education, 40,* 255-283.

Muffo, J. A. (1992). The status of student outcomes assessment at NASULGC member institutions. *Research in Higher Education, 33,* 765-774.

Ory, J. C., & Parker, S. A. (1989). Assessment activities at large, research universities. *Research in Higher Education, 30,* 375-385.

Patton, G. W., Dasher-Alston, R., Ratteray, O. M. T., & Kait, M. B. (1996). *Outcomes assessment in the Middle States Region: A report on the 1995 outcomes assessment survey.* Philadelphia, PA: Commission on Higher Education of the Middle States Association of Colleges and Schools.

Peterson, M. W., & Augustine, C. H. (2000a). External and internal influences on institutional approaches to student assessment: Accountability or improvement? *Research in Higher Education, 41,* 443–479.

Peterson, M. W., & Augustine, C. H. (2000b). Organizational practices enhancing the influence of student assessment information in academic decisions. *Research in Higher Education, 41,* 21–52.

Peterson, M. W., Augustine, C. H., Einarson, M. K., & Vaughan, D. S. (1999a). *Designing student assessment to strengthen institutional performance in associate of arts institutions.* Stanford, CA: Stanford University, National Center for Postsecondary Improvement.

Peterson, M. W., Augustine, C. H., Einarson, M. K., & Vaughan, D. S. (1999b). *Designing student assessment to strengthen institutional performance in baccalaureate institutions.* Stanford, CA: Stanford University, National Center for Postsecondary Improvement.

Peterson, M. W., & Einarson, M. K. (2000). An analytic framework of institutional support for student assessment. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 15). New York: Agathon Press.

Peterson, M. W., Einarson, M. K., Augustine, C. H., & Vaughan, D. S. (1999a). *Designing student assessment to strengthen institutional performance in comprehensive institutions.* Stanford, CA: Stanford University, National Center for Postsecondary Improvement.

Peterson, M. W., Einarson, M. K., Augustine, C. H., & Vaughan, D. S. (1999b). *Designing student assessment to strengthen institutional performance in doctoral and research institutions.* Stanford, CA: Stanford University, National Center for Postsecondary Improvement.

Peterson, M. W., Einarson, M. K., Augustine, C. H., & Vaughan, D. S. (1999c). *Institutional support for student assessment: Methodology and results of a national survey.* Stanford, CA: Stanford University, National Center for Postsecondary Improvement.

Peterson, M. W., Einarson, M. K., Trice, A. G., & Nichols, A. R. (1997). *Improving organizational and administrative support for student assessment: A review of the research literature.* Stanford, CA: Stanford University, National Center for Postsecondary Improvement.

Ratcliff, J. L., & Associates. (1995). *Realizing the potential: Improving postsecondary teaching, learning, and assessment.* University Park: The Pennsylvania State University, National Center on Postsecondary Teaching, Learning, and Assessment.

RiCharde, R. S., Olny, C. A., & Erwin, T. D. (1993). Cognitive and affective measures of student development. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 179–195). San Francisco: Jossey-Bass.

Rossman, J. E., & El-Khawas, E. (1987). *Thinking about assessment: Perspectives for presidents and chief academic officers.* Washington, DC: American Council on Education and the American Association for Higher Education.

Ryan, G. J. (1993). After accreditation: How to institutionalize outcomes-based assessment. In C. Praeger (Ed.), *Accreditation of the two-year college* (New Directions for Community Colleges No. 83, pp. 75-81). San Francisco: Jossey-Bass.

Sell, G. R. (1989). An organizational perspective for the effective practice of assessment. In P. J. Gray (Ed.), *Achieving assessment goals using evaluation techniques* (New Directions for Higher Education No. 67, pp. 21–41). San Francisco: Jossey-Bass.

Sims, S. J. (1992). *Student outcomes assessment: A historical review and guide to program development.* Westport, CT: Greenwood Press.

Steele, J. M., & Lutz, D. A. (1995). *Report of ACT's research on postsecondary assessment needs.* Iowa City, IA: American College Testing Program.

Steele, J. M., Malone, F. E., & Lutz, D. A. (1997). *Second report of ACT's research on postsecondary assessment needs.* Iowa City, IA: American College Testing Program.

Terenzini, P. T. (1989). Assessment with open eyes: Pitfalls in studying student outcomes. *Journal of Higher Education, 60,* 644–664.

Thomas, A. M. (1991). Consideration of the resources needed in an assessment program. *NCA Quarterly, 66,* 430–443.

Walleri, R. D., & Seybert, J. A. (1993). Demonstrating and enhancing community college effectiveness. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 87-102). San Francisco: Jossey-Bass.

Williford, A. M., & Moden, G. O. (1993). Using assessment to enhance quality. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 40–53). San Francisco: Jossey-Bass.

Wolff, R. A. (1992). CSU and assessment—second down and eight yards to go: A view from the scrimmage line, *Student outcomes assessment: What makes it work?* (pp. 73–80). Long Beach: California State University, Institute for Teaching and Learning.

Woodard, D. B., Jr., Hyman, R., von Destinon, M., & Jamison, A. (1991). Student affairs and outcomes assessment: A national survey. *NASPA Journal, 29*(1), 17-23.

Young, C. C., & Knight, M. E. (1993). Providing leadership for organizational change. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 25-39). San Francisco: Jossey-Bass.

Kenneth A. Feldman
Corinna A. Ethington
John C. Smart

A Further Investigation of Major Field and Person-Environment Fit

Sociological Versus Psychological Interpretations of Holland's Theory

In a recent article (Feldman, Smart, & Ethington, 1999), we used Holland's (1966, 1973, 1985, 1997) theory to study the extent to which the growth or achievement of college students—in terms of their gains in selected abilities and interests—was a function of the congruence or fit between these students' personalities and their major-field environments. The present analysis is an outgrowth and extension of that earlier analysis. It explores and compares the influence of personality dynamics and social-environmental forces on students' change and stability during college.

Holland's (1966, 1973, 1985, 1997) theory of personality types and environments assumes that most people (including college students) can be classified as one of six personality types (Realistic, Investigative, Artistic, Social, Enterprising, Conventional) based on their distinctive patterns of abilities, attitudes, and interests. The theory further provides three general propositions pertaining to college students and their development during their collegiate experience: (a) students search for and select academic environments that match their distinctive patterns of abilities, interests, and personality profiles; (b) academic environments

This article, an earlier version of which was presented at the annual meeting of the American Educational Research Association in New Orleans in April 2000, is partially based on materials found in Smart, Feldman, and Ethington (2000). We greatly appreciate the useful comments made by the three anonymous reviewers of this article.

Kenneth A. Feldman is a professor in the Department of Sociology, SUNY at Stony Brook; Corinna A. Ethington is a professor in the Department of Counseling, Educational Psychology, and Research, The University of Memphis; John C. Smart is a professor in the Department of Counseling, Educational Psychology and Research, The University of Memphis.

The Journal of Higher Education, Vol. 72, No. 6 (November/December 2001)
Copyright © 2001 by The Ohio State University

differentially socialize students toward the acquisition of distinctive patterns of abilities, interests, and values that reflect the abilities, interests, and values presumably reinforced and rewarded by the respective environments; and (c) student achievement is a function of the congruence or fit between their dominant personality type and the academic environment. Holland's theory is primarily psychological in origin and nature, and most research on the validity of the three propositions has focused on the first (self-selection) and third (congruence) propositions in efforts to assist students to select "proper academic majors" in which they stand the greatest likelihood of success and to better understand differential patterns of change and stability in students' abilities and interests during their undergraduate careers.

Abundant evidence generally supports the validity of each of the three propositions—especially those of self-selection and congruence (see, for example, the extensive literature reviews by Assouline & Meir, 1987; Holland, 1985, 1997; Spokane, 1985, 1996; Tranberg, Slane, & Ekeberg, 1993; Walsh & Holland, 1992); and reliance on Holland's theory as a theoretical framework for studies of college effects on students is growing (see, for example, Antony, 1998; Feldman et al., 1999; Huang and Healy, 1997; Thompson & Smart, 1999). In the present analysis we are not so much interested in offering additional evidence for any one of these three propositions as in assessing the relative merits of two of them—the second (socialization) and the third (congruence). The socialization and congruence propositions provide different—if not alternative—explanations for differential patterns of change and stability in students.

The socialization proposition postulates that the key element in promoting student acquisition of one rather than another set of competencies and talents is the academic environments (i.e., departments) students enter. Here, the roles of faculty members and their collective efforts to socialize students to the prevailing norms and values of their respective academic environments are the primary components to be considered, and the personality types and associated initial abilities and interests of students are of less importance and perhaps even irrelevant. That is to say, for example, that the likelihood of students collectively developing any specific repertoire of competencies and values is *singularly dependent* upon their entry into an academic environment that requires, reinforces, and rewards that particular repertoire. In this respect, the socialization proposition has a decided sociological orientation because of its focus on the collective group effects of academic environments. The effects of academic environments in Holland's theory are not assumed to be inherently dependent on the attributes of individual students who enter them. The respective academic environments are assumed to have similar effects on all students in them—irrespective of the students' personality types.

The congruence proposition assumes that—other things being equal—congruence of personality and environment is related to higher levels of educational stability, satisfaction, and achievement. Placing importance on congruence is intended to help individual students select educational settings in which they have a greater likelihood of long-term stability, satisfaction, and success. In principle, this emphasis on the individual in his or her environment reflects a social-psychological orientation underlying the congruence assumption, one which "blends" considerations of the personality type of students with the reinforcement efforts of faculty in the respective academic environments. According to the congruence assumption, the likelihood of a student developing any specific repertoire of competencies and values is *jointly dependent* on the student's own personality type and the congruence or fit between it and the student's entry into an academic environment that requires, reinforces, and rewards that particular repertoire. Thus, both the student's personality type *and* the substantive nature of academic environments are essential components in assisting individual students in the selection of educational settings in which they presumably have the greatest potential to persist, to be satisfied, and to be successful. Consideration of both the individual and the environment is seen as essential to understanding the potential consequences of individual behavior in academic settings. Although in principle the environment is as important as the individual's personality, in actuality, as Holland (1997) has noted, "the environmental models are only occasionally studied" (p. 60). Moreover, Walsh and Holland (1992) have concluded the following: "We view the theory as primarily psychological in nature and one in which personality variables are the most powerful and influential. . . . The theory tends to emphasize person variables and [is] lean on the concept of reinforcement [of the respective environments]" (p. 63). Consequently, in the present article we will refer to "personality dynamics" when interpreting certain findings relevant to congruence.

Our earlier analysis (Feldman et al., 1999) found considerable support for the congruence assumption for student change in Investigative, Artistic, and Enterprising abilities and interests but not for student change in Social abilities and interests. Students with dominant Investigative, Artistic, or Enterprising personalities who entered a major field congruent with their dominant personality gained (respectively) on Investigative, Artistic, or Enterprising abilities and interests, whereas those with these dominant personalities who did not enter a congruent major either remained the same or declined in these abilities. For example, the likelihood of students with a dominant Artistic personality type further developing their Artistic abilities and interests was contingent upon their entry into an Artistic academic environment: those who entered a congruent (Artistic) academic environment showed decidedly greater growth in these attributes than did those who entered an incongruent (non-Artistic) academic environment (where their Artistic abilities and interests actually declined in this particular case).

The data used for our earlier analysis, however, did not permit a comparison of the relative merits of person-environment congruence and the socialization influences of the environment in explaining student change and stability. We seek to accomplish this comparison in the present analysis by adding new data to the data already presented in the earlier analysis. In doing so, we hope to understand better the relative importance of psychological forces and social forces in their respective contributions to students' change and stability in college.

Research Procedures

Sample

As in our earlier study, data for the present study were obtained from the 1986 and 1990 surveys of the Cooperative Institutional Research Program (CIRP) sponsored by the Higher Education Research Institute at the University of California, Los Angeles. The overall sample consisted of 4,408 students attending 360 different institutions of higher learning. Students completed the standard CIRP freshman survey upon entering college in the fall of 1986 and a follow-up survey in the winter of 1990. The latter survey obtained information about the experiences of these students at college and how these students felt they had changed during the four years. Both the earlier and the present analyses are based on the responses of 2,309 students who were enrolled for all four years, whose academic major is included in Holland's (1997) classification of academic majors, and who provided complete information on all variables described below. Table 1 presents a listing of all majors classified into the four academic environments on which the present study focuses.

Variables

In the present study as in the earlier one, the classification of persons and of environments was done using the parallel typologies developed by Holland (see French, 1963, for a discussion of the importance of using such commensurate dimensions when measuring and studying person-environment fit).

TABLE 1

Academic Majors/Environments Represented in Student Sample

Realistic
Electrical Engineering
Mechanical Engineering
Marine Science
Drafting/Design
Military Science

Investigative
General Biology
Biochemistry/Biophysics
Botany
Marine (Life) Science
Microbiology/Bacteriology
Zoology
Other Biological Science
Finance
Aeronautical/Astronautical Engineering
Civil Engineering
Chemical Engineering
Astronomy
Atmospheric Science
Chemistry
Earth Science
Mathematics
Physics
Statistics
Other Physical Science
Pharmacy
Premedical/Predental/Preveterinary
Anthropology
Economics
Ethnic Studies
Geography
Sociology

Artistic
Arts
English
Language/Literature
Music
Speech
Theater/Drama
Music/Art Education
Architecture

Social
History
Philosophy
Theology/Religion
Elementary Education
Physical Education/Recreation
Special Education
Home Economics
Library Science
Nursing
Political Science
Psychology
Social Work
Women's Studies
Law Enforcement

Enterprising
Journalism
Business Administration
Marketing
Management
Business Education
Industrial Engineering
Communications
Computer Science

Conventional
Accounting
Secretarial Studies
Data Processing

Not in Holland
Other Humanities
Other Business
Secondary Education
Other Engineering
Health Technology
Therapy
Other Professional
Other Social Science
Other Technical
Agriculture
Other Fields

Classification of academic majors. The 1986 freshman survey asked students to select their "probable field of study," and the 1990 follow-up survey asked students to select their "current/last field of study" from a listing of academic disciplines/majors. For the earlier study, as well as this one, we classified these academic majors into the six academic environments proposed by Holland by using *The College Majors Finder* (Rosen, Holmberg, & Holland, 1989). A total of 64 of the 76 majors selected by the students could be thus classified. The Realistic and Conventional categories, however, had a combined total of only 4 academic majors with too few students to be useful in our analyses; consequently, these two categories were not included in either study. The number of students in each of the remaining four groups of academic majors is: Investigative ($n = 672$); Artistic ($n = 334$); Social ($n = 788$); and Enterprising ($n = 515$).

1986 and 1990 Ability and Interest Scales. The 1986 and 1990 CIRP surveys asked students to rate themselves compared with the average person their age on 12 different abilities (e.g., mathematical ability, social self-confidence) on a scale with 1 = lowest 10%, 2 = below average, 3 = average, 4 = above average, and 5 = highest 10%. Students were also asked to indicate the importance of 18 general goals and values (e.g., creating artistic work, being very well off financially) using a scale of 1 = not important, 2 = somewhat important, 3 = very important, and 4 = essential. From these items, we picked out those that were characteristic of adjectives used to describe the four Holland personality types considered in our analyses (Holland, 1997).

We used the twenty-six items we selected to create precollege (1986) and follow-up (1990) scales reflecting the distinctive abilities, interests, and goals that each of the four groups of academic majors, classified according to Holland's theory, are hypothesized to reinforce and reward. Except for a few items that required more extended discussion on our part, the items were easily classifiable because of Holland's extended analysis and description of each personality type in terms of a person's vocational and avocational preferences, life goals and values, self-beliefs, and problem-solving style (see Holland, 1997, chap. 2). To give one example, the CIRP general goal and value of "helping others who are in difficulty" unambiguously corresponds to the Social personality type, which Holland (1997) describes in part as a person who "perceives self as liking to help others" (p. 25) and "who wants to serve others in the context of medical support, institutional service or reciprocal interactions" (p. 25).

The scales used in this study, then, represent students' self-reported abilities and interests at the time they began college (1986 scores on the Investigative Artistic, Social and Enterprising scales) and four years later (1990 scores on these scales). Each of these scales (in both years) was created by standardizing the items and computing the average across items. Student scores were converted to T-scores with a mean of 50 and a standard deviation of 10. A complete listing of all items constituting each scale and the reliability estimates of all scales are presented in Table 2.[1] Note that our procedures produced acceptable reliabilities in the 0.60s and 0.70s.

Classification of personality types. Holland (1997) has noted that an individual's personality type may be measured by his or her responses to

TABLE 2

1986 and 1990 Student Ability and Interest Scales [a]

	Reliability	
1986 and 1990 Investigative Scale		
Self rating: Self-confidence (intellectual)	$\alpha = 0.682$	(1986)
Self-rating: Academic ability	$\alpha = 0.630$	(1990)
Self-rating: Mathematical ability		
Self-rating: Drive to achieve		
Goal: Making a theoretical contribution to science		
1986 and 1990 Artistic Scale		
Self-rating: Artistic ability	$\alpha = 0.683$	(1986)
Self-rating: Writing ability	$\alpha = 0.697$	(1990)
Goal: Becoming accomplished in one of the performing arts (acting, dancing, etc.)		
Goal: Writing original works (poems, novels, short stories, etc.)		
Goal: Creating artistic work (painting, sculpture, decorating, etc.)		
Goal: Developing a meaningful philosophy of life		
1986 and 1990 Social Scale		
Goal: Influencing the political structure	$\alpha = 0.750$	(1986)
Goal: Influencing social values	$\alpha = 0.794$	(1990)
Goal: Helping others who are in difficulty		
Goal: Becoming involved in programs to clean up the environment		
Goal: Participating in a community action program		
Goal: Helping to promote racial understanding		
1986 and 1990 Enterprising Scale		
Self-rating: Leadership ability	$\alpha = 0.752$	(1986)
Self-rating: Popularity	$\alpha = 0.762$	(1990)
Self-rating: Self-confidence (social)		
Goal: Become an authority in my field		
Goal: Obtaining recognition from my colleagues for contributions to my special field		
Goal: Having administrative responsibility for the work of others		
Goal: Being very well off financially		
Goal: Being successful in a business of my own		
Goal: Becoming an expert on finance and commerce		

Note: For the Self-rating items, students responded to the prompt "Rate yourself on each of the following traits as compared with the average person your age" using a scale of 1 = Lowest 10%; 2 = Below average; 3 = Average; 4 = Above average; 5 = Highest 10%. For the Goal items, students responded to the prompt "Indicate the importance to you personally of each of the following" using a scale of 1 = Not important; 2 = Somewhat important; 3 = Very important; 4 = Essential.

ability and interest scales. In both the earlier and the present study we therefore used the four 1986 ability and interest scales just described to determine each student's primary personality type. The profile for each of the 2,309 students was obtained and, in accordance with a suggestion by Holland (1997, p. 28), students were assigned to the personality type for which they had the highest scale score (in 1986).[2] This procedure resulted in the following distribution of students across the personality types: Investigative (*n* = 789); Artistic (*n* = 377); Social (*n* = 553); and Enterprising (*n* = 590).

Gender and consistency in choice of academic majors (type of recruit). In the earlier study, to trace possible conditional effects of gender and consistency or stability of decisions about academic major relative to Holland's propositions, we obtained information on the gender of students and whether their actual major in 1990 was the same or different from their intended major in 1986.[3] The total sample (n = 2,309) consisted of 1,217 female students (52.7%) and 1,092 male students (47.3%). Of the full sample, 1,128 students (48.9%) had similar majors/environments in 1986 and 1990 (and thus, in our terminology, were "primary recruits" to their 1990 major) and 1,181 students (51.1%) had dissimilar or different academic majors/environments over the four-year period (and thus were "secondary recruits" to their 1990 major).

Analyses

The earlier analysis. Our analytical approach in the earlier analysis relied on a pretest, posttest longitudinal design with repeated measures, which uses students as their own controls. We adapted a procedure developed by Roberts (1980) to adjust initial scores for regression-to-the-mean bias (for details see Feldman et al., 1999, pp. 649–650). These adjusted scores were used in all analyses that examined patterns of change among students between 1986 and 1990. We defined "higher levels of achievement" in Holland's theory as students' positive gain ("growth") in abilities and interests reflective of their dominant personality type. Similarities and differences among students were examined at two points in time (1986 and 1990) to assess the proposition that congruence between dominant personality of students and their academic environments fosters higher levels of achievement of the distinctive abilities and interests characterizing the four personality types and academic environments. We used repeated measures analysis of variance to examine this assumption. Separate analyses were conducted for each of the four ability and interest scales described in Table 2. For each scale, the analyses used only those subjects whose dominant personality matched the particular scale. For example, in examining changes in Enterprising abilities and interests, the sample used were those students whose dominant personality type was Enterprising. Results of these analyses are given in Appendix A.

The within-subjects independent variable in the repeated measures analysis of variance was the student's self-rated assessment of abilities and interests in 1986 and 1990. Three between-subjects factors were included in the analysis to test for the possibility of interactions between overall change and (1) congruence between the student's personality type and his or her actual (1990) academic environment, (2) gender of the student, and (3) similarity or dissimilarity of the student's intended academic environment in 1986 and actual academic environment in 1990 (as an indicator of whether the student was a primary or secondary recruit to the environment). Our interest in these analyses was the pattern of change and stability in the self-rated interests and abilities of students over time based on the congruence between their personality type and their actual academic environment as well as the extent to which these changes were similar for men and women and for primary and secondary recruits.

Each of the four analyses had two levels for each of the three between-subjects factors. The two gender levels were males and females. The two levels in the similarity of students' 1986 and 1990 academic majors were students having similar majors (primary recruits to the 1990 major) and dissimilar majors (secondary recruits to the 1990 major) at these two points in time. The two levels of congruence distinguished between students who majored in academic environments that were congruent or incongruent with their dominant personality type. A significant interaction between students' congruence and their academic environment and year was considered as indicating that gains on the initially prominent abilities and interests between 1986 and 1990 depended on their selection of a congruent or incongruent academic environment. When significant interactive effects were found for gender or type of recruit, the congruence-by-year interactions were plotted separately for the subgroups of students.

As already noted, our earlier analysis basically supported the congruence assumption (as conceived in that analysis) for student change in Investigative, Artistic, and Enterprising abilities and interests but not for student change in Social abilities and interests. Those students with dominant Investigative, Artistic, or Enterprising personalities who entered a major field congruent with their dominant personality gained (respectively) in Investigative, Artistic, or Enterprising abilities and interests, whereas those with these dominant personalities who did not enter a congruent major either remained the same or declined in these abilities. An interaction effect of congruence and gender was not found for any of the abilities and interests, and so the data for male students and female students were combined rather than presented separately. Only one interaction of congruence with type of recruit emerged (for Investigative abilities and interests), thereby necessitating separate consideration of primary and secondary recruits when looking at their changes in these abilities and interests.

The present analysis. In the present analysis we explore the patterns of change and stability of students' abilities and interests of students whose personality types were congruent and of those who were incongruent with a *given* environment to determine if the effects of a particular environment were comparable for the two different groups of students. At first glance, it might seem that this comparison is the one we already made in our earlier analysis, but it is not. In the earlier analysis we explored the change and stability of students on traits associated with an environment whose personality type was congruent with that environment but *not* the change and stability of students whose personality types were incongruent with that particular environment. Rather, the comparison group for a particular personality type (say an Artistic personality type) in a congruent environment (Artistic environment) was that particular personality type (Artistic) in a noncongruent environment (in this case, Investigative, Social, or Enterprising environments). We now conceive of this latter group of students (already studied in our earlier analysis) as evidencing *one kind of incongruence.* Not considered in this earlier analysis was the change and stability of students in a *particular* academic field (say, the Artistic field) who were not congruent with *that particular field* (that is, students who were Investigative, Social, and Enterprising personality types in the Artistic field). Data for this *second type of incongruence* form the core of the present analysis.

We find it convenient to refer to two kinds of incongruence—Incongruent I and Incongruent II—as given in the pertinent tables and figures of the present analysis. Technically speaking, however, it is not so much that two different kinds of incongruence are being defined, but that two different sets of classifications of the full set of incongruent students are being considered. In our earlier study, students in the (full) set of incongruent students were classified by their dominant personality (in order to study their change or stability in the abilities and interests reflective of their dominant personality type). In the present study, the same students in the (full) set of incongruent students are reclassified by the academic environments of which they became a part (in order to study their change or stability in the abilities and interests promoted by these particular environments).

Data for this second type of incongruence are given in the present article in Tables 3 and 4 and Figures 1 through 5. We have produced much of these data by adding information to the data already presented in the earlier analysis (see Table 3 and Figures 1 through 5 in Feldman et al., 1999). In fact, Table 3 of the present article is the same as the parallel table in the earlier article (coincidentally also numbered Table 3)—with the important exception that the third and sixth column of Table 3 of the present article are new (and to make this new information stand out, we have put it in italics). Likewise, the five figures of the present article are the same as the earlier ones except that a third graph line (non-bold, non-broken) has been added to each of the figures. It is these additional data in the tables and figures that will help us distinguish between the effects of a socialization dynamic and a personality dynamic. If students whose personality type is congruent with a particular environment consistently make greater gains than do students whose personality type is incongruent with that particular environment, then one might say a "personality dynamic" is at work, in that being the "right personality type" enhances gain over and above the influence of the environment. On the other hand, if the pattern of change is the same for the two groups of students, then we might say that a "socialization dynamic" is at work because even a "wrong personality type" in a particular environment shows a pattern of change similar to that of the "right personality type."

In our earlier analysis, it was found that (with one exception) growth in a particular area was dependent on the congruence of major and personality. That is, there were interactions between environment and year, and the plots of those interactions showed that the lines indicating change for congruent students and incongruent students (the first type of incongruence) were not parallel. In now examining whether a "personality dynamic" or "socialization dynamic" is at work, the focus is on whether there is an interaction of personality type and year. If there is no interaction, the growth pattern (and thus the slopes of the lines indicating change) would be the same for students whose personality was congruent with the environment and those whose personality was incongruent (the second type of incongruence); thus the growth patterns would be parallel and a socialization dynamic would be suggested. By contrast, a significant interaction would mean that the patterns of change were different for the two groups of students and that the lines indicating change were not parallel, thereby suggesting the presence of a personality dynamic.

Repeated measures analysis of variance was used to assess changes in the four 1986 and 1990 ability and interests scales between college entry (in 1986) and four years later. Because our earlier analysis had tested for interactive effects by gender and type of recruit, the present analysis focused only on the student's personality type. Again, separate analyses were conducted for each of the four abilities and interests scales. Because the earlier analyses had found interactive effects due to type of recruit in the analyses of the Investigative scales, the present analyses were conducted separately for the primary and secondary recruits into Investigative environments. For each analysis there was one between-subjects factor (personality type) and one within-subjects factor (year).

Results

Table 3 presents the means on the four 1986 and 1990 ability and interest scales for students classified according to the congruence (or lack thereof) between their personality type and academic environment. These means are plotted in Figures 1 through 5 as well. The magnitudes and direction of change for students are also shown on the graphs in Figures 1 through 5 in terms of effect sizes. These effect sizes, which were calculated using the error mean squares from the within-subject analyses, convert the difference between the 1986 and 1990 means for each group to standard-deviation units. The other sets of effect sizes shown in the figures were calculated using the error mean squares from the between-subject analyses; they indicate the differences between students in the different environments (congruent and incongruent) in 1986 as well as in 1990.

We are particularly interested in the result for congruent and incongruent students in the *same* major-field environment, for this is the comparison that could not be made in our earlier analysis given the particular data set used in that analysis. We focus initially, then, on only Columns 1, 3, 4, and 6 in Table 3 and only the two non-broken lines (one bolded, one not) in each of the figures. This focus further entails consideration of data in Table 4, which presents the results of the repeated measures analyses of variance for the two groups of students under consideration. Note that in each instance in Table 4 the interaction between

TABLE 3

Means and Standard Deviations for 1986 and 1990 Abilities and Interests Scales

	1986			1990		
	Congruent	Incongruent I	*Incongruent II*	Congruent	Incongruent I	*Incongruent II*
Investigative Scale						
Primary recruits	57.209	52.576	*50.732*	58.737	53.079	*53.032*
	(5.228)	(4.736)	*(4.999)*	(9.065)	(9.001)	*(8.911)*
Secondary recruits	55.109	54.155	*48.712*	57.201	52.285	*50.465*
	(4.945)	(5.354)	*(5.648)*	(8.705)	(8.863)	*(9.626)*
Artistic Scale[a]	58.580	56.293	*53.743*	62.369	54.872	*58.107*
	(5.429)	(6.591)	*(6.814)*	(9.782)	(10.208)	*(10.138)*
Social Scale[a]	53.798	54.275	*48.810*	54.609	53.730	*50.133*
	(4.106)	(4.243)	*(4.003)*	(9.824)	(9.880)	*(9.865)*
Enterprising Scale[a]	55.187	55.241	*48.629*	57.394	55.312	*51.011*
	(4.860)	(4.887)	*(5.206)*	(9.057)	(9.480)	*(9.638)*

NOTE: Data in Columns 1, 2, 4, and 5 originally appeared in Table 3 in Feldman, Smart and Ethington (1999) while data in Columns 3 and 6 (in italics) are new to this article; standard deviations are in parentheses; see text and figures for definitions of the two types of incongruence.

[a] Results are similar for males and females, primary and secondary recruits.

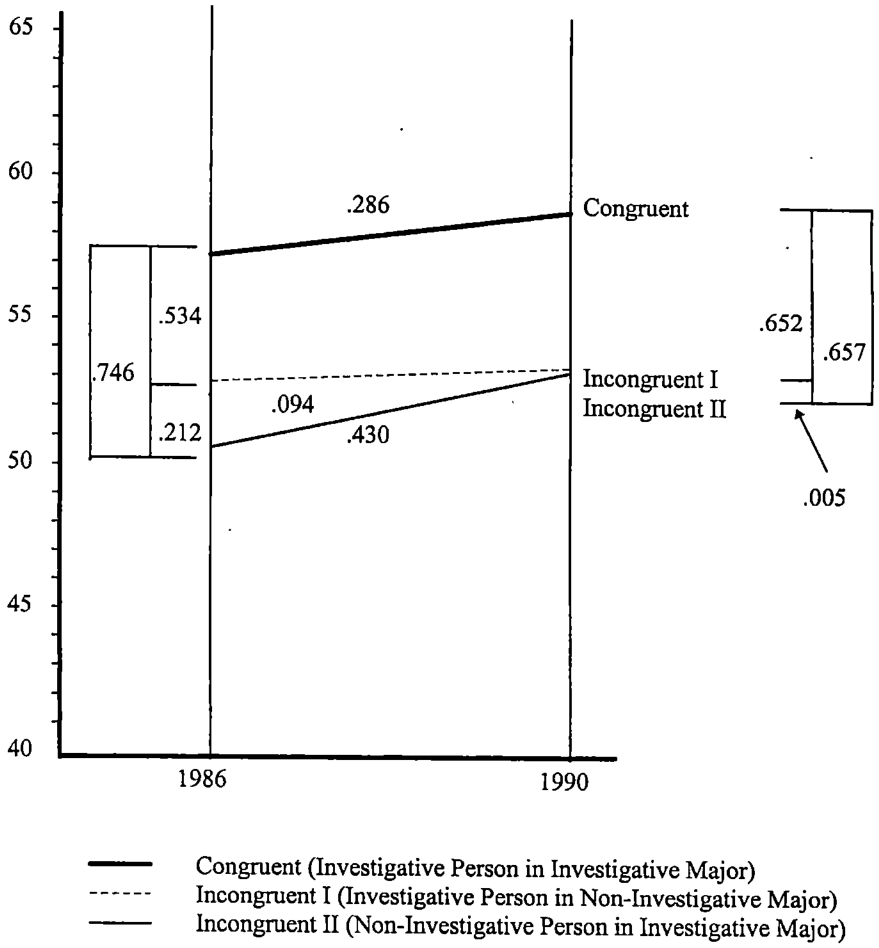


FIG. 1. Patterns of Change in Investigative Abilities and Interests–Primary Recruits

personality type and year was *not* significant, indicating that the direction and magnitude of change between 1986 and 1990 was similar for students whose personalities were congruent with the environment and those whose personalities were incongruent with the environment. The significant effect of personality type indicated that the two groups differed in their abilities and interests. This, together with the nonsignificant interaction, indicated that differences between the two groups of students were essentially the same in 1986 and in 1990.

Taken together as a set, the two tables and five figures not only show certain similarities across the four academic environments but also reveal a socialization dynamic rather than a psychological dynamic at work. Thus, in each of the five figures the pattern for the second type of

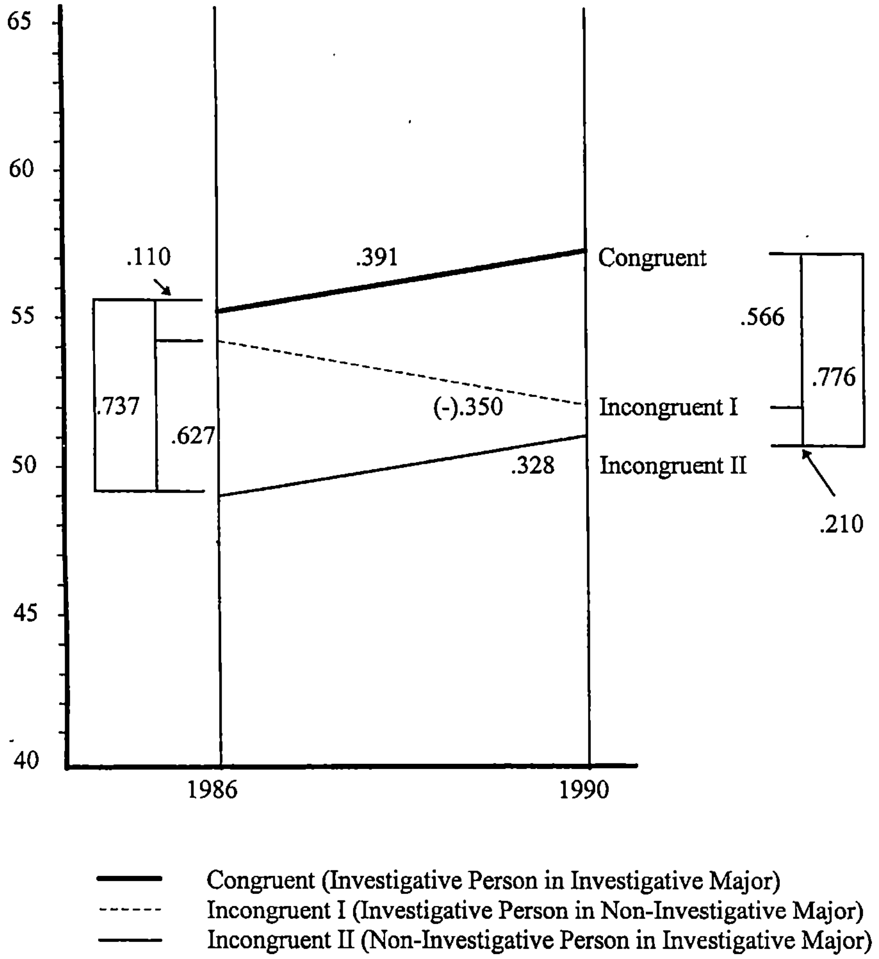


FIG. 2. Patterns of Change in Investigative Abilities and Interests–Secondary Recruits

incongruent student (non-bold, non-broken line) and the congruent students (bold, non-broken line) are, for all intents and purposes, *parallel*—confirmed by the statistical analyses presented in Table 4. That is, the pattern of gain or growth on the analogous cluster of abilities and interests for each academic environment is essentially similar for all students in that environment, whether or not the student's dominant personality type is congruent or incongruent with that environment. Although students whose personality type is incongruent with the environment begin with a lower score in 1986, they experience basically the same amount of gain as those whose personality type is congruent with the environment.

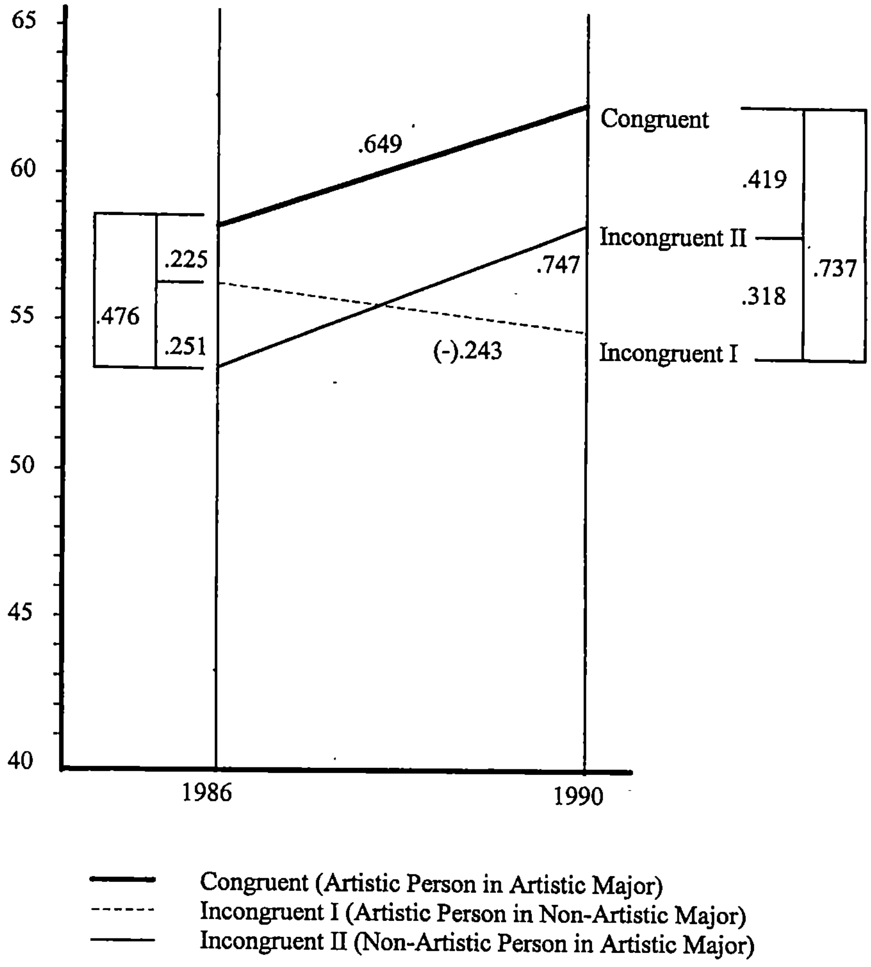


FIG. 3. Patterns of Change in Artistic Abilities and Interests

To focus on one case in point, consider the pattern of change in Enterprising abilities and interests of students in Enterprising fields (shown in Table 3 and Figure 5). Students whose dominant personality was congruent with such fields (that is, students who had an Enterprising personality type) had a mean scale score of 55.187 in 1986 and one of 57.394 in 1990, thereby showing an average gain of 2.207 (effect size = 0.381). Students in the same field who did not have an Enterprising personality type also gained in mean scale score—from 48.629 in 1986 to 51.011 in 1990, thus showing an average gain of 2.382 (effect size = 0.411). This same pattern of parallel gains can be seen in the other major fields (including both primary and secondary recruits in Investigative

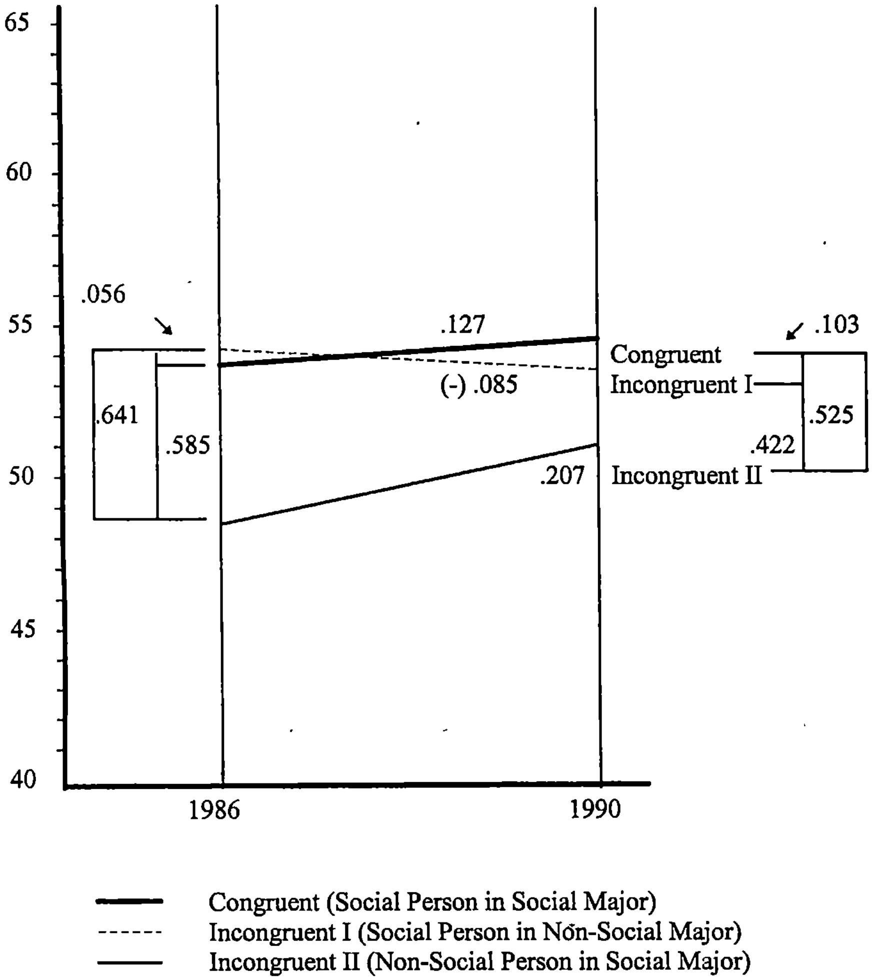


FIG. 4. Patterns of Change in Social Abilities and Interests

fields). This particular pattern clearly suggests that an essentially uniform socialization dynamic is at work within each of the four academic environments rather than a personality dynamic. Of interest here is that what we take to be a socialization effect on the incongruent students was not so strong as to close (or significantly reduce) the initial gap between them and the congruent students (thus parallel lines rather than converging lines).[4]

Of some additional interest is a comparison of the two types of incongruent students (see the non-bold, broken line and the non-bold, nonbroken line in each figure). Notice for each environment that the second

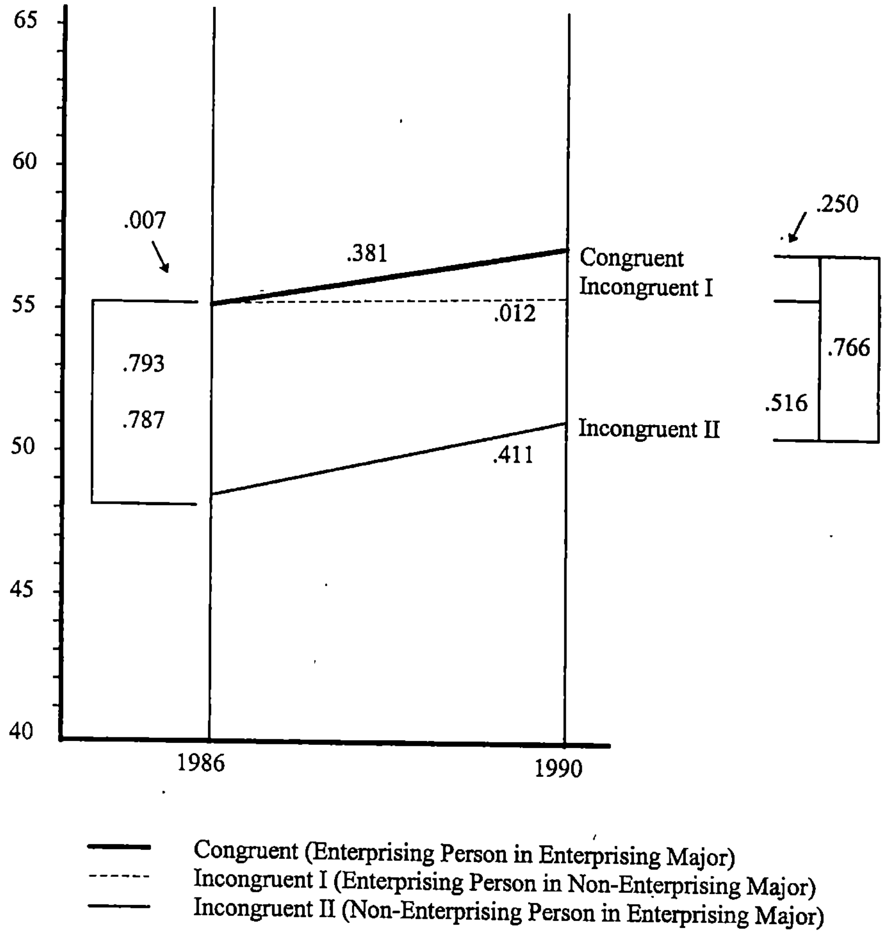


FIG. 5. Patters of Change in Enterprising Abilities and Interests

type of incongruent student gained on average between 1986 and 1990 in the abilities and interests being considered, whereas the first type of incongruent student—as shown in our earlier analysis and repeated in the present analysis—either remained about the same or actually decreased in these abilities and interests. The largest gain (in terms of effect sizes) on a given cluster of abilities and interests by the second type of incongruent student was made by students in the Artistic environment (0.747), followed by primary recruits in Investigative environments (0.430) and students in the Enterprising environment (0.411), and finally by secondary recruits in Investigative environments (0.328) and students in Social environments (0.207). In short, students who did *not* enter a "fitting" environment (first type of incongruence) either maintained or

TABLE 4

Analysis of Variance Results: 1986 and 1990 Abilities and Interests Scales

Source	Investigative (Primary Recruits) ($df = 1,363$)		Investigative (Secondary Recruits) ($df = 1,305$)		Artistic ($df = 1,332$)		Social ($df = 1,786$)		Enterprising ($df = 1,513$)	
	MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between Subjects Effects										
Personality Type	6145.22	78.40**	5901.78	70.81**	3081.39	28.05**	7904.56	107.86**	10082.64	121.11**
Within plus Residual	78.38		83.35		109.84		73.28		83.25	
Within Subject Effects										
Year	606.82	20.68**	505.87	15.29**	2474.04	78.98**	401.84	10.03*	1268.04	40.38**
Personality Type by Year	24.69	0.84	3.92	0.12	12.32	0.39	23.23	0.58	1.84	0.06
Within plus Residual	29.35		33.08		31.33		40.08		31.41	

$*p < 0.01$. $**p < 0.001$.

decreased in the interests and abilities of the environment they did not enter, whereas students who entered a "nonfitting" environment (second type of incongruence) actually increased in the interests and abilities of the environment they did enter.

It is also instructive to note that, in all cases, the second type of incongruent student started with a lower mean score in 1986 than did the first type of incongruent student. For instance, the mean score on Artistic abilities and interests for the second type of incongruent student (53.743 in Table 3) is lower than the mean score for the first type of incongruent student (56.293 in Table 3). But in all instances—due we speculate to the greater prevalence of a socialization dynamic than a personality dynamic—the initial gap between these two groups shrunk by 1990. This is most evident for the findings concerning the longitudinal pattern of scores for Artistic abilities and interests (Figure 3) where the second type of incongruent students actually ended up with a higher mean score than the first type of incongruent students.

Finally, we call attention to one other interesting aspect of the results. Despite the positive increases in abilities and interests evidenced by the second type of incongruent students, which paralleled the positive increases made by congruent students, in no instance were the final (1990) scores of the second type of incongruent students as high as the entering (1986) scores of the congruent students on the pertinent cluster of abilities and interests under consideration. The closest scores to being so were those by the Incongruent II students in Artistic fields, whose 1990 average score of 58.107 on the Artistic Abilities and Interests Scale was almost as high as the 1986 average score of 58.580 of the congruent students on the same scale.

Discussion

Comparing the Findings of the Earlier Study and the Present Study

It might be thought that the major finding of the present study—that students whose dominant personality type was *incongruent* with an academic environment made about the same amount of gain in abilities and interests pertinent to that environment as did students whose dominant personality type was *congruent* with that environment—is somehow contradictory with the results of our earlier analysis. We do not see this as being so and offer the following explanation. Our earlier analysis focused specifically on particular subsets of students based upon their dominant personality type and examined their longitudinal pattern of change and stability on a specific cluster of abilities and interests that paralleled (or fit) their personality type. Those findings (repeated in the pertinent tables and figures of this article) did in fact support the congruence assumption in that—for instance—students with a dominant Artistic personality tended to grow or gain in Artistic abilities and interests if they entered Artistic, but not other, academic environments. In this sense, congruent academic environments reinforced the dominant pattern of abilities and interests of this particular subset of students. If the question is the extent to which academic environments are tools to perpetuate the distinctive abilities and interests of students at the time they enter college, then the answer is, "yes they are": those distinctive patterns of abilities and interests will be enhanced if, and only if, students enter academic environments that reinforce and reward those specific abilities and interests. Indeed, as seen in the pertinent tables and figures of the present article, such students ended up higher on these particular abilities and interests than did either of the other two comparison sets of students under consideration.

But now a different question in terms of the potential impacts of academic environments can be raised: to what extent are academic environments successful in their efforts to socialize a disparate collection of students to the distinctive pattern of preferred abilities and interests of the environments? Are these environments as "effective" with students who begin with lower levels of commensurate abilities and interests as those with higher levels of commensurate abilities and interests? The various profiles of patterns of students evidencing the second type of incongruence reported in this article provide the basis for an affirmative response: that is, yes, the impacts of academic environments appear to be comparable for students whose personality types are congruent or incongruent with the respective environments (although the initial gap between the two groups of students does not lessen). Thus, the two sets of findings are not in conflict, but rather are directed toward two distinct but related questions.

The potential for confusion in predictions and interpretations arises because the positive effects of congruence found in the comparisons made in our first and earlier analysis need not of necessity be found in the comparisons made in this, our second, analysis. Had they in fact been found in this second analysis, the effects of the respective academic environments would have been greater for students with congruent personality types than for those students with incongruent personalities. But this is not what the data of the present analysis showed.

Holland's theory provides essentially no guidance in terms of the relative merits of a personality perspective and a socialization perspective (and any predictions that would follow therefrom) because the theory does not seek to disentangle potential group and individual effects. From the *group* perspective, the socialization proposition implicitly postulates a uniform pattern of reinforcement and reward by faculty members in the respective academic environments, or at the very least, does not consider potentially different patterns of longitudinal change and stability in student abilities and interests depending on students' congruence with the environment. This is so because the concern of the socialization assumption is on the collective actions and effects of academic environments. From the *individual* perspective, the congruence proposition postulates a differential pattern of longitudinal change and stability in abilities and interests for comparable individuals (i.e., two students with similar personality profiles): those entering congruent environments will grow or gain more than those entering incongruent environments. But the congruence assumption is silent as to the collective effects of the respective academic environments on students in them with different personality profiles—presumably because the concern of the congruence assumption is the pattern of change and stability of individual students within the *different* environments. In the present analysis, then, we were unable to make formal predictions about the possibility of uniform versus differential patterns of change in student abilities and interests in the *same* academic environment.

Limitations of the Present Study and Some Directions for Future Research

As is inevitably true, the use of secondary analysis limits what can and cannot be researched and analyzed. Thus, although our study had data from students at a wide variety of colleges, it was limited to analyzing primarily full-time students who basically spent four continuous years at the same college and most of whom were about to graduate. We were not able to study the importance of academic environments to students who attend more than one college during their college careers or students who spread their college education over five or more years (some of whom may be in part-time attendance). Further research would be needed to see if the results of our study would also pertain for these other sorts of students.

The use of CIRP data also limited the way we could measure personality types. We had to construct abilities and interest scales from the items available in the CIRP survey, which were then used to determine a student's dominant personality type. The more common way within the Holland tradition to assess a person's personality type is to use the person's scores on established personality or interest inventories such as the

Self-Directed Search (SDS; Holland, Fritzsche, & Powell, 1994), the Vocational Preference Inventory (VPI; Holland, 1966), the Strong-Campbell Interest Inventory (SCII; Campbell & Hansen, 1981), and the Strong Vocational Interest Blank (SVIB; Campbell & Hansen, 1981). Specifically, the theme scores of the SCII, the composite activities, competencies, occupations, and self-rating scales from the SDS, and the occupational preference scales of the VPI may be used to assess a person's resemblance to the types proposed by Holland. Whether the use of one of these inventories would produce more precise or different results than we obtained can be determined only by future research.

We should note that even with the data available to us, we could have classified students and their academic environments in more complex ways than we did. For example, instead of classifying the student by his or her dominant personality type alone (using his or her highest scale score among the abilities and interest scales), it would have been possible for us to determine each student's personality *profile* or *pattern* across two or more of these scales (for details, see Holland, 1997, chap. 2) Likewise, there are ways of differentiating various environmental profiles or patterns (Holland, 1997, chap. 3). Thus congruence/incongruence can be measured by taking into account more than the dominant personality type of the students and more than the simplest or least undifferentiating coding of the environment. The way we determined congruence and incongruence (either Incongruent I or Incongruent II), has been denoted as the dichotomous first-letter agreement index by Brown and Gore (1994), and is only one of a number of ways that have been used in the research literature (and which we could have used with the data available to us). Brown and Gore compare nine other congruence indices as to their underlying distribution characteristics and their abilities to discriminate among subtle but important differences in person-environment congruence. The benefits of using one or another of these indices with the present research design and data (or one like it) await further research.

To make a final observation in this section, we first remind the reader that the central and most important finding of the present analysis was that the congruent and incongruent students in the same academic environments made parallel gains in interests and abilities during their college years. This is not to say, however, that there might not be important differences between the two groups of students that deserve to be researched. Despite the incongruent (II) students in a particular academic environment making the same amount of change as the congruent group of students within a particular academic environment, the absolute scores of this second kind of incongruent students were always lower than those of the respective congruent students (both at the beginning and near the end of college); moreover, the absolute scores of the second kind of incongruent students were generally more like those of the first kind of incongruent students than they were like the congruent students (especially near the end of college). Could it be, then, that despite showing parallel gains, the congruent and incongruent students within academic environments are distinctive in their perceptions of the college experience, in their satisfactions and dissatisfactions with the college and major-field environments, in the ease with which they do change at college and the personal "costs" involved, and the like? To determine whether this is so, more detailed survey information would be needed than we had at our command. Moreover, we would imagine that the comparison of these two groups of students using in-depth *qualitative* research would be of particular benefit in understanding the impacts of academic environments on their students.

The Importance of Academic Environments

The centrality of academic environments in understanding differential patterns of change and stability in students' abilities and interests discerned from our analyses parallels the findings of Pace (1990) that academic majors are an important influence "on the extent and direction of students' progress in college" (p. 76) and his conclusion that "students learn what they study" (p. 76). Some might interpret this conclusion narrowly as implying that academic majors and environments are simply "knowledge factories" dispensing only subject matter knowledge and content. Such an interpretation would be consistent with Pascarella and Terenzini's (1991) conclusion that "students tend to demonstrate the highest levels of learning on subject-matter tests most congruent with their academic major . . . [whereas] the impact of major field of student on noncognitive outcomes is substantially less apparent or consistent" (p. 614).

We would argue, however, that academic environments are something more than merely "knowledge factories" dispensing straightforward subject-matter knowledge and content. We would suggest the following interpretation based on the premises of Holland's theory and our own findings. Holland's (1997) theory postulates that environments require, reinforce, and reward a wide range of abilities, interests, attitudes, and values. His descriptions of the six environments repeatedly note that they foster the development of different competencies and achievements, encourage people to see themselves and the world in different ways, stimulate people to engage in different kinds of activities, and reward people for their display of alternative values and attitudes (see pp. 43–48). Theoretically, then, academic environments do not merely dispense subject-matter knowledge and content. Rather, they seek to influence a full spectrum of personal and professional self-perceptions, competencies, attitudes, interests, and values.

Our findings similarly show that academic environments contributed to gain or growth in a range of student abilities, interests, and attitudes reflected in the diversity of items and scales used to assess change and stability of students—for the scales on which gains were found included a relatively broad range of measures of students' self-reported abilities (e.g., academic, artistic, leadership, mathematical, writing), self-concepts (e.g., intellectual and social self-confidence, popularity), and interests (e.g., making a theoretical contribution to science, writing original works, helping to promote racial understanding, being very well off financially). We were limited, of course, to the use of the items included in the CIRP survey and thus were not able to include in our study an even wider set of abilities, interests and attitudes. Nevertheless, the range of available items would seem broad enough to be able to conclude that thorough knowledge of academic disciplines and environments is a prerequisite to understanding differences in the patterns of student change and stability resulting from the college experience.[5]

As we noted previously, although work and academic environments are an essential component in Holland's theory, the research literature on such environments *per se* is not as large as that on other aspects of the theory (Holland, 1997; Walsh & Holland, 1992). Gottfredson and Richards (1999) concluded, for instance, that although Holland's theory has stimulated extensive research, the vast majority of studies have been focused on the hypothesized behaviors of individuals (derived from Holland's theory) rather than on the collective actions and effects of work and academic environments. This is unfortunate, because Holland (1997) has characterized his theory as a structural-interactive or typological-interactive theory, noting that "it is structural or typological because it attempts to organize the vast sea of information about people *and* jobs . . . [and] interactive because it assumes that many careers and social behaviors are the outcome of people *and* environments acting on one another" (p. 12, emphasis added). The findings of the present study and earlier research (Antony, 1998; Huang & Healy, 1997; Smart, 1997; Smart & Feldman, 1998), provide strong support for the socialization proposition in Holland's theory that environments are a major influence on the educational and vocational behaviors of their members. So do still earlier studies, such as Walsh and Lacy (1969, 1970) and Walsh, Vaudrin and Hummel (1972), whose findings have been interpreted as showing the "molding effects" of environments (see Walsh & Holland, 1992, pp. 60–61). The findings from these various and scattered studies, then, suggest the need for even greater as well as more systematic attention to the collective actions and effects of environments in subsequent research on Holland's theory.

APPENDIX A

Analysis of Variance Results of 1986 and 1990 (Self-Reported) Abilities and Interests Scales

Source	Investigative ($df = 1{,}781$)		Artistic ($df = 1{,}369$)		Social ($df = 1{,}545$)		Enterprising ($df = 1{,}582$)	
	MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between Subjects Effects								
Holland environment	4653.80	61.75***	3932.83	38.04***	35.69	0.49	334.38	4.81*
Type of recruit	238.47	3.16	33.84	0.33	46.35	0.64	11.10	0.16
Gender	1776.48	23.57***	1034.11	10.00**	3.67	0.05	3030.52	43.60***
Holland by type of recruit	363.12	4.82*	18.76	0.18	148.28	2.04	90.87	1.31
Holland by gender	99.71	1.32	0.93	0.01	172.38	2.37	238.50	3.43
Type of recruit by gender	92.58	1.23	230.52	2.23	12.95	0.18	91.81	1.32
Holland by type of recruit by gender	232.87	3.09	85.30	0.83	215.84	2.97	72.29	1.04
Within plus residual	58863.68		38146.07		39651.61		40452.77	
Within Subject Effects								
Year	74.74	2.62	260.52	7.63**	6.19	0.15	255.52	7.62**
Holland by year	401.36	14.05***	956.14	28.02***	85.25	2.08	317.96	9.49**
Type of recruit by year	68.64	2.40	5.17	0.15	1.59	0.04	3.26	0.10
Gender by year	245.68	8.60**	92.11	2.70	204.23	4.98*	913.26	27.25***
Holland by type of recruit by year	215.48	7.54**	11.95	0.35	144.78	3.53	0.10	0.00
Holland by gender by year	65.31	2.29	74.91	2.20	16.89	0.41	0.28	0.01
Type of recruit by gender by year	1.16	0.04	0.03	0.00	27.20	0.66	50.60	1.51
Holland by type of recruit by gender by year	15.53	0.54	1.77	0.05	47.11	1.15	6.36	0.19
Within plus residual	22314.63		12591.21		22331.76		19508.54	

SOURCE: This table originally appeared as Appendix A in Feldman, Smart, & Ethington (1999)
$*p < 0.05$. $**p < 0.01$. $***p < 0.001$.

Notes

[1] We emphasize that the items constituting these scales are based on students' self-reports of their abilities and interests. Pike (1995, 1996) has presented considerable evidence showing the relationships between students' self-reports and objective measures of abilities based on the use of confirmatory factor analysis and multitrait-multimethod analytical procedures. Although he cautioned that there is a need for further research on the relationships between content correspondence and convergence before self-reported data can be fully accepted as "valid proxies of policy indicators" (Pike, 1995, p. 19), he concluded that "using self-reports as general indicators of achievement can be justified" because "self-reports and test scores based on the same set of specifications do represent the same educational outcome domains (i.e., they are congeneric)" (Pike, 1996, p. 110).

[2] A student was classified as a particular personality type even if his or her highest scale score was only one or two points higher than on another of the scales. We experimented with setting a higher "bar" for classification as a particular personality type—for example, using a criterion of five points higher on one scale than on any other scale. Using more stringent criteria did not seem to make much difference in preliminary analyses of our data, however, so we kept our approach of using the highest scale score (no matter how big or small) of a person to represent his or her dominant personality.

[3] Gender differences had emerged in both Feldman and Weiler's (1976) study and Smart and Feldman's (1998) study of accentuation effects of academic departments. Feldman and Weiler also presented indirect evidence that longitudinal patterns of change and stability in students' personality profile scores and self-description indices differed for students with similar and dissimilar initial and final academic majors.

[4] Our interpretation of these gains as socialization effects is based not only on the data of the present analysis but also on several other sets of data to be found in our full report (see Smart, Feldman, & Ethington, 2000).

[5] Our findings clearly have many practical implications for affecting the educational experiences of college students— including advising students in picking courses and selecting academic majors, enhancing student persistence both in the freshman year and the later years of college, and promoting a more systematic broadening of students' learning experiences. These and other educational and policy implications of the particular findings of this article, as well as the more general findings of our larger report, can be found in Chapter 9 of Smart, Feldman and Ethington (2000).

References

Antony, J. S. (1998). Personality-career fit and freshman medical career aspirations: A test of Holland's theory. *Research in Higher Education, 39*, 679–698.

Assouline, M., & Meir, E. I. (1987). Meta-analysis of the relationship between congruence and well-being measures. *Journal of Vocational Behavior, 31*, 319–332.

Brown, S. D., & Gore, P. A., Jr. (1994). An evaluation of interest congruence: Distribution characteristics and measurement properties. *Journal of Vocational Behavior, 45*, 310–327.

Campbell, D. P., & Hansen, J. C. (1981). *Manual for the SVIG-SCII.* Stanford, CA: Stanford University Press.

Feldman, K. A., Smart, J. C., & Ethington, C. A. (1999). Major field and person-environment fit: Using Holland's theory to study change and stability of college students. *Journal of Higher Education, 70*, 642–669.

Feldman, K. A., & Weiler, J. (1976). Change in initial differences among major-field groups: An exploration of the "accentuation effect." In W. H. Sewell, R. M. Hauser, &

D. L. Featherman (Eds.), *Schooling and achievement in American society* (pp. 373–407). New York: Academic Press.

French, J. R. P., Jr. (1963). The social environment and mental health. *Journal of Social Issues, 19*(4), 39–56.

Gottfredson, L. S., & Richards, J. M., Jr. (1999). The meaning and measurement of environments in Holland's theory. *Journal of Vocational Behavior, 55,* 57-73.

Holland, J. L. (1966). *The psychology of vocational choice.* Waltham, MA: Blaisdell.

Holland, J. L. (1973). *Making vocational choices.* Englewood Cliffs, NJ: Prentice-Hall.

Holland J. L. (1985). *Making vocational choices* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Holland J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Holland, J. L., Fritzsche, B. A., & Powell, A. B. (1994). *The self-directed search technical manual.* Odessa, FL: Psychological Assessment Resources.

Huang, Y., & Healy, C. C. (1997). The relations of Holland-typed majors to students' freshman and senior work values. *Research in Higher Education, 38,* 455-477.

Pace, C. R. (1990). *The undergraduates.* Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles.

Pascarella, E. G., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Pike, G. R. (1995). The relationship between self-reports of college experiences and achievement test scores. *Research in Higher Education, 36,* 1–21.

Pike, G. R. (1996). Limitations in using students' self-reports of academic development as proxies for traditional achievement measures. *Research in Higher Education, 37,* 89–114.

Roberts, A. O. H. (1980). Regression toward the mean and the regression-effect bias. *New Directions for Testing and Measurement, 8,* 59–82.

Rosen, D., Holmberg, K., & Holland, J. L. (1989). *The college majors finder.* Odessa, FL: Psychological Assessment Resources.

Smart, J. C. (1997). Academic subenvironments and different patterns of self-perceived growth during college: A test of Holland's theory. *Journal of College Student Development, 38,* 68-77.

Smart, J. C. & Feldman, K. A. (1998). "Accentuation effects" of dissimilar academic departments: An application and exploration of Holland's theory. *Research in Higher Education, 39,* 385-418.

Smart, J. C., Feldman, K. A., & Ethington, C. A. (2000). *Academic disciplines: Holland's theory and the study of college students and faculty.* Nashville, TN: Vanderbilt University Press.

Spokane, A. R. (1985). A review of research on person-environment congruence in Holland's theory of careers. *Journal of Vocational Behavior, 26,* 306–343.

Spokane, A. R. (1996). Holland's theory. In D. Brown & L. Brooks (Eds.), *Career choice and development* (3rd ed., pp. 33–74). San Francisco: Jossey-Bass.

Thompson, M. D., & Smart, J. C. (1999). Student competencies emphasized by faculty in disparate academic environments. *Journal of College Student Development, 40,* 365-376.

Tranberg, M., Slane, S., & Ekeberg, E. (1993). The relation between interest congruence and satisfaction: A meta-analysis. *Journal of Vocational Behavior, 42,* 253–264.

Walsh, W. B., & Holland, J. L. (1992). A theory of personality types and work environments. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person-environment psychology: Models and perspectives* (pp. 35-69). Hillsdale, NJ: Lawrence Erlbaum Associates.

Walsh, W. B., & Lacy, D. W. (1969). Perceived change and Holland's theory. *Journal of Counseling Psychology, 16,* 348-352.

Walsh, W. B., & Lacy, D. W. (1970). Further exploration of perceived change and Holland's theory. *Journal of Counseling Psychology, 17,* 189–190.

Walsh, W. B., Vaudrin, D. M., & Hummel, R. A. (1972). The accentuation effect and Holland's theory. *Journal of Vocational Behavior, 2,* 77-85.

Nira Hativa
Rachel Barak
Etty Simhi

Exemplary University Teachers

Knowledge and Beliefs Regarding Effective Teaching Dimensions and Strategies

Rationale and Theoretical Framework

The quest for excellence in college and university teaching is now a worldwide concern. Universities pay increasing attention to the quality of the pedagogy practiced in their classrooms and to assessing how effectively professors are teaching (Ovando, 1989). At the same time, educators and researchers are looking for ways to increase knowledge about teaching effectiveness. A good way to achieve this goal is to learn from outstanding teachers about how they think about teaching, about their pedagogical knowledge, and about their instructional behaviors in order to convey all these to less successful teachers.

Teachers' Thinking, Beliefs, and Knowledge About Teaching

We refer to teacher thinking and knowledge interchangeably because separating these two notions is difficult, if not unfeasible (Clark & Peterson, 1986). Of all components of teacher thinking, we are particularly interested in teachers' beliefs. The need to learn about teachers' thinking, beliefs, and knowledge results from an emerging image of the teacher as a "thoughtful professional." Teachers' thought processes, pedagogical knowledge, and beliefs were found to substantially affect their classroom behavior at the precollege level (Clark & Peterson, 1986; Fennema & Franke, 1992; Peterson, 1988; Thompson, 1992). Effective teaching requires a wide knowledge base that can be broken down into

Nira Hativa is a professor and Head, University Center for Staff Development in Education, Tel Aviv University. Rachel Barak and Etty Simhi were graduate students when this article was written, School of Education, Tel Aviv University.

The Journal of Higher Education, Vol. 72, No. 6 (November/December 2001)
Copyright © 2001 by The Ohio State University

several main categories (Shulman, 1987; Wilson, Shulman, & Richert, 1987). We concentrate here on general pedagogical knowledge defined as "knowledge of, and skill in, the use of teaching methods and pedagogical strategies that are not subject-specific" (Wilson et al., 1987, p. 114).

University professors, not having received any systematic preparation for their teaching role, gain beliefs and knowledge about good pedagogy through trial-and-error in their work, reflection on student feedback, and by using self-evaluation. To a much lesser extent, they learn from having observed their own teachers while they were students (Hativa, 1997). This unplanned and nonsystematic process may lead to fragmented pedagogical knowledge and to unfounded beliefs about what makes teaching effective. It is interesting to learn about the pedagogy-related thinking, beliefs, and knowledge, particularly of outstanding teachers, because these probably contribute to their excellent instruction.

Exemplary Teachers' Teaching-Related Thinking, Beliefs, and Knowledge

Research at the precollege level provides considerable evidence that exemplary teachers (sometimes referred to as expert teachers) differ from their colleagues, and particularly from novice teachers, in the complexity and sophistication of their thought about teaching; in their cognitive schemata and pedagogical reasoning skills (Borko & Livingston, 1989); in their decision making (Westerman, 1991); and in their teaching-related knowledge. They integrate the different categories and forms of knowledge in ways that allow them to optimally structure the physical, social, and intellectual environment of their classrooms (Gudmundsdottir & Shulman, 1987).

There is similar evidence at the college level. Exemplary teachers, usually selected on the basis of high student ratings, were found to prefer a deep approach to teaching, which is incongruent with students' more common surface approach (Andrews, Garrison, & Magnusson, 1996). Compared with other teachers, they have a more extensive, complex, and flexible repertoire of concepts of teaching effectiveness, they hold more developed concepts of self-efficacy, they use a wider range of criteria for self-evaluation, and they draw upon almost twice as many strategies for enhancing student learning. In addition, exemplary teachers were more inclined to believe that they played a significant role in their students' learning, and they were significantly more confident that they possessed the teaching skills necessary to give effect to their potential (Dunkin, 1995; Dunkin & Precians, 1992). These studies suggest that exemplary teachers have a more developed pedagogical knowledge than other teachers. However, we do not know whether this knowledge is fully developed or whether the same pedagogical knowledge is shared by all exemplary teachers.

Exemplary Teachers' Characteristics and Classroom Behaviors

At the precollege level, characteristics and classroom behaviors of expert teachers were found to differ from those of novice teachers in several ways. Expert teachers presented multiple points of view and a wider perspective; they showed more active, sensitive, and deliberative involvement; their problem-framing strategy was different, their assessment of students' nonverbal behavior was more accurate (Ropo, 1987; Stader, Colyer, & Berliner, 1990), and their response, in terms of instructional interventions, to students' behavior was more successful (Ropo, 1987).

Studies at the college level found that exemplary teachers are highly organized, plan their lessons carefully, set unambiguous goals, and have high expectations of their students. They give students regular feedback regarding their progress in the course, make specific remediation recommendations, and assume a major responsibility for student outcomes. They make course content relevant to students by giving examples and connecting course goals to the expectations and experiences of their students (Horan, 1991). They treat students as individuals in the classroom, encourage them, challenge them intellectually, use a variety of teaching strategies to add interest, and create a positive classroom environment (Hilgemann & Blodget, 1991). Exemplary teachers enjoy teaching, show enthusiasm for the subject, have excellent command of the language and good delivery, inject humor, and introduce dramatic elements. They make an earnest attempt to promote students' learning, and actively involve them in the learning process through questions and discussions (Kelly & Kelly, 1982). They offer clear, organized, and interesting presentations, and they communicate positive regard to students and motivate them (Lowman, 1996). A review of observational studies on effective classroom teaching behaviors (Murray, 1997) found enthusiasm/expressiveness, clarity of explanation, and rapport/interaction to be the main effective teaching characteristics. A review of experimental studies (Murray, 1997) suggests that classroom teaching behaviors in the enthusiasm and clarity domains appear to be causal antecedents (rather than mere correlates) of various instructional outcome measures. In sum, exemplary university teachers are well prepared and organized, present the material clearly, stimulate students' interest, engagement, and motivation in studying the material through their enthusiasm/

expressiveness, have positive rapport with students, show high expectations of them, encourage them, and generally maintain a positive classroom environment.

These findings are supported by correlational studies on effective college teaching. A review of such studies (Feldman, 1989a) found that among the 22 main teaching characteristics examined, "Clarity and understandableness" and "Teacher preparation and organization" are the most important teaching characteristics as perceived by teachers and students. Next in importance were "Stimulation of interest," "Motivation of students to reach high standards," "Encouragement of discussion and openness to others' opinions," and "Elocutionary skills." Interestingly, this review indicates that students and faculty closely agree on the definition of effective teaching characteristics. Centra found an even wider scope of agreement: "Colleagues, administrators, students, and alumni have all identified similar general characteristics of what effective teachers do" (1996, P. 55).

However, all these studies leave open the question whether some teaching characteristics are more important than others, that is, whether they are prerequisite for exemplary instruction. Other questions are, what particular strategies do exemplary teachers use in class—do they all use the same strategies to achieve their excellence? If not, what is the interplay between the various possible strategies? Are some strategies used more frequently than others? Is there a core of strategies that are critical and indispensable for achieving effective teaching? This study attempts to answer these questions in the limited context of several case studies.

Relationships Between Teachers' Thinking, Beliefs and Knowledge, and Classroom Practice

As presented above, studies at the precollege level suggest that teachers' thinking, pedagogical knowledge, and beliefs regarding teaching and students substantially affect their classroom behavior. However, these studies found inconsistent relationships between teachers' thinking and knowledge and their observed practice. This inconsistency suggests that teachers' thinking and knowledge are not related in a simple causal way to their instructional practices but rather that the relationship is complex, with many factors influencing teacher work, such as the social context in which teaching takes place, including the values, beliefs, and expectations of peers, academic administrators, and students (Thompson, 1992).

The meager research on this topic at the college/university level suggests the same inconsistent relationships between teachers' thinking and knowledge and their observed practice. Hativa (2000c) identified a substantial discrepancy between faculty's perceptions of their most important teaching goals, or their beliefs regarding how successful they are in achieving these goals, on one hand, and students' perceptions of this success on the other. Students felt that faculty failed to achieve by far the large majority of the goals regarding teaching effectiveness. While faculty perceived themselves as satisfactory teachers who had sufficient general pedagogical knowledge and who applied it well in practice, students were strongly critical of the level of teaching in their courses. They felt that faculty failed in their classroom practice—that the teaching they experienced was unclear, boring, and generally ineffective. There is strong supporting evidence for the gap between faculty and students' perceptions of effectiveness of university teaching. Mostly the inconsistency is between faculty who consider themselves good teachers and their students who think otherwise (Berman & Skeff, 1988; Blackburn, Boberg, O'Connell, & Pellino, 1980; Cross, 1977; Feldman, 1989b).

Only few studies suggest some explanation for this gap. Three case studies of university professors whose students perceived as poor teachers (Hativa, 1998; in print) revealed that they all suffered from insufficient general pedagogical knowledge and that they failed to apply most of the effective strategies, even though they were familiar with them. McAlpine and Weston (in print) investigated teaching and reflection on teaching of six expert teachers and then compared them with studies of nonexpert teachers. They found that moving from better thinking about teaching to better enactment of teaching requires linking knowledge and experience to action through reflection, practice and feedback over time. They concluded that better teaching knowledge does not necessarily lead to better teaching.

Aims of the Study

This study concentrates on general pedagogical knowledge that is not subject matter related and involves pedagogical principles and classroom strategies/behaviors that were found in research to contribute to effective instruction. Exemplary teachers were the objects of study because the research literature indicated that their thinking about teaching and their general pedagogical knowledge were more developed than those of other teachers, as in the case of their performance on effective teaching strategies.

Thus, the main aim of our study was to identify the beliefs and general pedagogical knowledge of exemplary university teachers regarding effective teaching strategies, the extent to which they use various of these strategies, and the relations between their beliefs and knowledge to their classroom practice. The research questions were: 1. What are their beliefs and general pedagogical knowledge about effective teaching dimensions and classroom strategies? 2. What effective teaching strategies do they actually use in class? 3. What are the relationships between their beliefs and knowledge about effective teaching strategies and their actual use of these strategies?

Method

To study teacher thinking and knowledge we used qualitative methods. The data were gathered in a research university in Israel during the second semester of the 1998 academic year. The *teachers studied* were four exemplary teachers—two from each of two different departments—Hebrew literature and psychology. These departments were selected because the undergraduate degrees of two of the researchers in this study were respectively in these two academic areas, so that they had content knowledge of the subject matters. In each department we listed all instructors whom students had rated high on teaching large undergraduate courses (at least 50 students) during the three years preceding the study. Then we selected from this list all instructors who were scheduled to teach a large undergraduate course in the semester of the study. From this final list (of three and four professors, respectively) we selected two instructors per department, and secured their consent to participate in the study. During the 14-week semester, each course met for three weekly academic hours.

Teacher A, in his late thirties, had been teaching for ten years in the literature department. His elective course, attended by 55 students, dealt with satire in the theatre. *Teacher B,* a woman in her early fifties, had been teaching in the same department for over twenty years. Her course, attended by 127 students, dealt with travel stories in Hebrew literature. *Teacher C* from the department of psychology was a woman in her early forties. She had joined the department four years earlier after serving for five years as a faculty member in an Australian university. Her course dealt with psychological testing, was required for students majoring in psychology, and 125 students were registered for it. *Teacher D* was a man in his late fifties, a veteran in the psychology department. He had an international reputation as a specialist in the psychology of terror and participated in trials of terrorists worldwide, was frequently invited to participate in TV debates on terrorism, and also taught courses on this topic in US universities. His course on terrorism and political violence was elective and was attended by 110 students from Psychology and other departments.

The instructional unit under study was selected from the course curriculum by each teacher before the beginning of the course. This was a self-contained unit of instruction that lasted from three to four consecutive class meetings.

The judges were the second and third authors of this article. They were trained by the first author on evaluating effective classroom teaching behaviors from videotaped lessons.

Data Collected

To achieve triangulation of the findings, we used several data sources, as follows.

Teachers' interviews. Each teacher was interviewed twice. Each interview was tape-recorded:

(a) A semistructured precourse interview, which took place a few days before the beginning of the semester. It dealt with all study questions, with more questions added during the interview, building on the teacher's responses.

(b) An open post-unit interview, which took place several days after completion of the instructional unit selected for study. The professor was asked to reflect on his/her teaching, on whether or not she or he achieved instructional goals, and on problems or unexpected events encountered while teaching.

Students' interviews. Ten students in each course under study were randomly selected for interviews about the instructor's classroom behavior.

Videotaped classes. The first lesson of the semester and all lessons related to the instructional unit under study were fully videotaped.

Effective-Teaching questionnaire. At the end of the last lesson of the instructional unit under study, a specially designed questionnaire was administered to all students in class. The questionnaire form asked to evaluate the four *main dimensions of effective teaching,* as identified in the literature presented above: lesson organization, clarity, interest/student engagement, and positive classroom climate. We included four additional *general items:* a global item about "Teacher overall performance," and three items found in the literature to be important components of effective teaching and often included in instructor rating forms: "Preparedness for lesson," "Intellectual challenge and promoting self thinking," and "Effective use of class time." In addition, for each of the four main dimensions we included several classroom strategies that are components of that dimension, taken from a structured model of dimensions and strategies of effective teaching (Hativa, 1999; 2000a; 2000b). To illustrate, the strategy "Listing the lesson outline on the board at the lesson beginning" contributes to the main dimension of "Lesson organization." Altogether, the questionnaire consisted of 36 Likert-type items.

The questionnaire was pilot tested in two courses in the psychology and literature departments, other than the courses designated for study. Participants were 81 students, and the results of both courses were combined. Coefficients of internal consistency (Cronbach's alpha) were computed. The contribution of each item to its content-wise respective scale was examined, and two items whose scale reliability had a detrimental effect were removed—one from the "organization" scale, and the other from the "interesting" scale. The scale of "classroom environment" failed to yield homogeneous clusters, and therefore its four items were also removed. Thus, the final questionnaire form included 30 items, 22 of which presented classroom behaviors/strategies. The reliability of the subscales was found to be: organization (5 items) 0.59, clarity (13 items) 0.69, and interest (4 items) 0.72.

Materials distributed to students by the instructor. For each course, the syllabus, other materials distributed to the students by the instructor, and the test forms were gathered.

Data Analysis and Results

All taped interviews and videotaped lessons were transcribed and the content analyzed. The categories in the content analysis were developed to represent the issues of this study. The initial categories were based on the 36 original items of the questionnaire, but in the course of analysis some of these categories were omitted and others that emerged from the data were added. Thus, at the end of this process the items rated in the interview and videotape analysis coincided only partially with those of the questionnaire. The coding of classroom teaching behaviors from videotape showed high levels of inter-rater reliability, as supported by Murray (1997).

First Question: What Are Exemplary Teachers' Beliefs and General Pedagogical Knowledge About Effective Teaching Dimensions and Strategies?

The answer to this question is based on the analysis of the instructors' pre- and post-interviews, in which they were asked to describe their beliefs and knowledge about dimensions of effective teaching and about effective teaching strategies and to describe those strategies that they used in class. The instructors were then asked specifically what strategies they used to make their lessons organized, clear, interesting, and inducive of a positive climate in class. Finally, they were presented with a list of strategies and asked about their use of these particular strategies in teaching (Table 1).

Teachers' thinking and knowledge on effective teaching strategies. Excerpts 1–4 represent, in short, some of the teachers' most interesting expressions of their thinking and beliefs about the main dimensions of effective teaching, taken from the transcripts of their interviews.

> *Excerpt 1: Teachers' Beliefs and Knowledge Concerning Lesson Organization*
> Teacher B opens each travel story by presenting some hypothesis, theory or axiom, and then she proves the hypothesis. The proof develops through analysis of the story and relates to all parts of the story. On the other hand, this hypothesis can be seen to emerge from all the story details. Thus, there is a cyclic process—the argument builds on the details, but the details prove the argument. Probably for this reason she calls herself very disorganized. *Teacher C* implements the structure of introduction, body, and conclusion on the topic of study rather than on the full lesson. "I start a topic, e.g., the Rohrschach test, with historical background on its development, then I describe the test with examples and illustrations, and then I let students experience some versions of the test for themselves. I end up by putting this test under the general framework of the category of tests it belongs to."
>
> *Excerpt 2: Teachers' Beliefs and Knowledge Concerning Lesson Clarity*
> *Teacher A* provides illustrations he judges relevant to the students—politics, general literature, psychology, etc. He speaks slowly enough to make it comfortable to absorb what he says. *Teacher B* believes that language is the main tool of the lecturer and that she achieves clarity by using clear diction and rich language, but also simplification and spiral repetition and enhancement of the ideas. She encourages students to offer their own examples and adapts her explanations to the two types of audiences attending her course—regular Israeli students and new immigrants who are non-fluent in Hebrew. For example, she assigns special reading material to those with insufficient command of Hebrew. *Teacher C* separates what is important from what is not, in order to weed out too many details that may obstruct understanding of the main points. *Teacher D* checks students' command of the background knowledge necessary to learn the new material and adapts his teaching style and level to them.
>
> *Excerpt 3: Teachers' Beliefs and Knowledge Concerning Interest in Lesson*
> *Teacher A* believes in the need to select material that is interesting and engaging to the students, and thus he presents entertaining excerpts from satires. He introduces variety to the lesson by using movies, videotapes, and

humorous anecdotes; inviting guest speakers; or even going out with the students to see a play. He assigns reading of interpretive or controversial articles for intellectual stimulation and for promoting knowledge in domains unfamiliar to students. He moves in front of the class and never reads from his notes. *Teacher B* believes that what makes a lesson interesting is a combination of the instructor's personality, method of teaching, and presentation of interesting/attractive topics. She strongly believes that attending class should give students some added value beyond what they can read by themselves. She enjoys teaching and communicates this to her students. To achieve interest she selects attractive topics, jokes a lot in her lessons, and integrates texts from other domains such as history, social sciences, psychology, and other cultures. She uses maps, paintings, or picture books to illustrate points. Occasionally she even takes her students to a movie or a play that is relevant to the topics taught. *Teacher C* uses thinking challenges to increase curiosity and lead to interaction with students. If a student asks an interesting question that is relevant to the topic, she opens it for discussion in order to stimulate students' thinking and to activate them. She lets students experience personality testing by assigning them to take the test and then grade themselves on a scale she provides. She believes this motivates students because it personalizes the material and makes it more lively and interesting for them. She also uses transparencies to illustrate and demonstrate the different tests. *Teacher D* presents thinking challenges in the form of questions that increase curiosity and tension and intellectually activate students during the lesson. He maintains interactions with students through discussion and participation. He uses spontaneous humor and shows enthusiasm in presentation. "I very much like the material I'm teaching—I live it all the time, and I think that I communicate this enthusiasm to students—I hope I cause them to get enthusiastic about it too."

Excerpt 4: Teachers' Beliefs and Knowledge Concerning Achieving a Positive Classroom Climate

Teacher A believes that his attitude toward students should be very positive and patient. He exhibits willingness to listen to and interact with them, communicates with them in and out of class, and prevents competition, which he perceives as damaging to classroom climate. *Teacher B* believes that her main characteristic as a teacher is friendliness toward students. She believes that teaching should be open and avoid inducing any feelings of anxiety. She helps students who face difficulties, particularly when they do not know sufficient Hebrew to follow the lesson. She communicates to students the feeling that they have interesting things to study and that with a lot of work, time and patience, they will succeed. *Teacher C* exhibits care for and interest in the students and listens to them throughout the lesson. She talks with students as if she were one of them—they joke and laugh together. She makes herself very approachable to her students—encourages them to ask questions, to talk to her either before or after the lesson, or to come to her office even outside the official office hours. *Teacher D* emphasizes the importance of presenting a variety of viewpoints rather than imposing his own. He believes that by this he promotes students' openness to others' opinions and pluralism. He avoids being critical or judgmental toward students or insulting them because this might create antagonism towards the teacher and it

> "shuts students' ears." He is very accessible to students and enables them to talk to him before or after class or during office hours.

The following summary is based on the full transcripts of the interviews, not only the excerpts presented above.

Lesson organization. While Teacher A emphasizes the major importance of organizing the lesson, and Teacher C perceives it as important, the other two teachers do not think it is important to make the lesson organized. Only Teacher A knows and uses the well-known lesson organization strategies of three-part division, outlining, dividing into topics and subtopics, and summarizing. Teacher B explicitly declares herself to be disorganized, and the remaining two teachers mention only a few organization strategies. Interestingly, two teachers (B and C) take the topic—the satire/story/test—as a unit of organization rather than the lesson.

Lesson clarity. All four teachers recognize the importance of clear teaching, of posing questions during the lesson to check for student understanding, and of giving good examples and illustrations. Three of them mention the need to link the examples to students' prior knowledge or experience, and three (not the same ones as before) simplify their explanations. All other strategies are mentioned by two teachers at most.

Making a lesson interesting/engaging. All four teachers recognize the importance of maintaining student attention throughout the lesson. They say they provide motivation to study the material and use teaching aids. All four also introduce variety into the lesson through changing teaching methods, using teaching aids, inviting guest speakers, and so forth. Three of the four teachers mention enthusiastic and dynamic presentation, challenging students' thinking, using humor, and activating students during the lesson.

Classroom climate. All teachers are very much aware of the need to create a pleasant classroom climate that is conducive to learning. They all exhibit care and respect for students. Three of them provide encouraging feedback and make themselves very accessible to students.

Table 1 compares the three sources—teachers' interviews, classroom videotapes, and students' rating—on the level of use of classroom strategies by the four teachers. The list of 34 strategies is a combination of the items in the questionnaire and those identified from content analysis of the interviews and videotapes, as already described. Strategies that appeared only in a single source were omitted from the table to enable comparison of at least two sources. The marks in the table are explained below.

TABLE 1

Effective Classroom Techniques/Strategies—Low Inference Teaching Behaviors: Evaluation from Instructor Interviews, Classroom Videotapes, and Students' Ratings

Source:	Instructors' Interviews				Classroom Videotapes				Modified Students' Ratings (Table 2)			
Department:	Literature		Psychology		Literature		Psychology		Literature		Psychology	
Instructor:	A	B	C	D	A	B	C	D	A	B	C	D
LESSON ORGANIZATION/Mean per instructor:					4.5	2.8	4.3	3.0				
Links lesson to overall course framework	v		v		5	4	5	3	c	d	e	c
Links lesson to previous lesson	v			v	5	4	4	5				
Presents lesson outline at lesson beginning	v			v	4	1	4	3	a	a	d	d
Summarizes the lesson at end	v				3	1	2	2	b	d	b	a
Divides lesson/topic into parts/topics/theses	v		v		5	5	5	3	c	d	e	c
Marks moves from one topic to another	v				5	2	5	2				
LESSON CLARITY/Mean per instructor:					3.8	4.0	4.2	3.8				
Simplifies explanations		v	v	v	5	5	5	4	d	d	e	d
Presents questions to check understanding	v	v	v	v	4	3	3	2	a	a	c	b
Encourages students' questions			v	v	5	3	5	5				
Answers well students' questions		v			3	4	5	5	c	c	e	e
Emphasizes important points	v		v		5	5	4	3	d	e	d	d
Writes (legibly) on board/transparencies					1	1	5	3	a	a	e	c
Provides intuitive meaning to explanations					2	5	3	4	b	d	c	d
Gives appropriate examples/illustrations	v	v	v	v	4	5	5	5	c	e	e	e
Links to students' prior knowledge/experience	v	v		v	3	4	2	1	a	d	c	a
Repeats/elaborates difficult points	v	v			4	4	4	5				
Speaks intelligibly	v	v			5	4	4	5	d	d	d	e
Explains unfamiliar/difficult terms					4	4	5	3	b	d	e	d
Does not make errors/mistakes					5	5	5	5	e	e	e	e

TABLE 1 (*Continued*)

Source:	Instructors' Interviews				Classroom Videotapes				Modified Students' Ratings (Table 2)			
Department:	Literature	Psychology			Literature		Psychology		Literature		Psychology	
Instructor:	A	B	C	D	A	B	C	D	A	B	C	D
INTERESTING/ENGAGING LESSON/Mean per instructor:					3.8	4.4	3.5	4.6				
Provides motivation to study	v	v	v	v	4	5	4	5	b	d	d	e
Presents enthusiastically/dynamically		v		v	3	5	3	5	b	e	d	e
Uses analogies, metaphors, anecdotes	v	v		v	3	5	2	5	b	e	c	e
Provides philosophical/historical background				v	4	5	3	4	b	e	c	d
Presents thinking challenges	v		v	v	4	5	5	5				
Uses teaching/technological aids	v	v	v	v	5	2	5	3				
Introduces variety into lessons	v	v	v	v	5	4	4	5				
Refers to own life or professional experience	v			v	2	5	2	5				
Integrates material from other domains	v	v			2	4	3	5				
Uses humor	v	v		v	5	5	2	5				
Activates students during the lesson	v		v	v	5	3	5	4				
CLASSROOM CLIMATE/Mean per instructor:					4.8	5.0	5.0	5.0				
Exhibits care for students	v	v	v	v	5	5	5	5				
Behaves respectfully towards students	v	v	v	v	5	5	5	5				
Provides encouraging feedback to students	v		v	v	5	5	5	5				
Approachable to students	v		v	v	4	5	5	5				

NOTES: 1 = *very low* through 5 *very high*.
A rule of thumb (Hativa, 1995) roughly interprets the mean ratings on a scale from 1 to 5 as follows: (a) below 2.9 very low; (b) 3.0–3.4 low; (c) 3.5–3.9 medium; (d) 4.0–4.4 as high; and (e) 4.5–5.0 very high.
[1] Also belongs to the category of Clarity.

The first four columns on the left present those strategies that teachers were familiar with, as identified from their interviews. The two judges went over all the interview transcripts and checked each behavior teachers mentioned during the interviews on the list (a yes/no rating). The judges agreed on over 90% of the listed behaviors and only those were marked by a "v" in Table 1.

Table 1 reveals substantial variation among the teachers regarding knowledge of the 34 strategies. Teachers A and D mention the largest number of strategies (25 and 21, 74% and 62%, respectively), whereas Teachers B and C mention only 16 (47%) strategies each. Strategies of organization seem to be the least known to our teachers, and only seven (21%) of the 34 strategies are familiar to all four teachers.

Second Question: What Strategies Do Exemplary Teachers Actually Use in Class?

The answer is based on both the analysis of the videotaped classes and students' mean ratings of their teachers. The following is a discussion of each source separately.

Source: analysis of classroom videotapes. Table 1 (middle four columns) presents evaluation of classroom behaviors of the teachers as analyzed from the videotaped lessons. This evaluation is expressed in numbers from 1 through 5, assigned to the items, using the following procedure. Each judge watched the tapes separately and rated the quality of the listed behaviors on a Likert scale from 1 = very low through 5 = very high. To illustrate, each of the judges evaluated the quality of "Speaks intelligibly" for the four teachers as either high (4) or very high (5), that is, both judges perceived the speech of all teachers as intelligible or even very intelligible. Several of the items enabled frequency counts, for example, "Gives good examples and illustrations." The judges first counted the number of occurrence in all videotaped lessons for a teacher and then computed the mean frequency of their occurrence per lesson. Then they worked out a Likert scale for the frequency of each item. For example, on the latter item, the worked out scale was: 1 (very low) for giving no example; 2 (low) for giving 1–2 examples; 3 (medium) for giving 3–5 examples; 4 (high) for giving 6–9 examples, and 5 (very high) for giving 10 or more examples. Teacher A gave, on average, 6 examples per lesson, which was rated 4 by the judges, whereas the other three teachers gave, on average, between 18 and 26 examples during a lesson, which judges rated as 5.

Inter-judges agreement was soundly high. Both judges assigned the same number for about 25% of the items rated. For about 40% of the other items, the gap in ratings between the two judges was of one whole number (e.g., 3 and 4) and the higher number (4) was chosen to represent the item evaluation. For about 25% of the items, the gap was of two numbers (e.g., 3 and 5), and in this case the mean of the two numbers (4) was chosen. For the remaining 10% of the items for which the gap between both raters was larger than two points, the judges discussed each others' ratings and agreed on the assigned number.

Excerpts 5-8 below present a short summary drawn from the two judges' written descriptions of the videotaped lessons. These descriptions provide further explanations and support to the numerical ratings in the middle columns in Table 1.

Excerpt 5: Analysis of Classroom Videotapes of Teacher A

Teacher A uses many strategies of lesson organization and clarity. In the first lesson of the course he draws a timeline chart representing the course plan, and at the beginning of each subsequent lesson he marks where they are on the chart and connects the lesson to the previous lesson. All lessons have the same structure: they start with a short presentation of a satire, the reason why he has selected it, historical background, and then the main content. He builds the material from explicit categories and subcategories and marks each transition between them. He supports each explanation by at least one example and frequently uses the technique of rule-example-rule (first presenting some idea or thesis, then providing an example to support it, and then repeating the general rule). He poses questions to students to check their understanding and repeats the explanation if necessary. He is not dynamic or enthusiastic in presentation and gains students' attention through the use of humorous video clips that demonstrate the points he makes and through the use of spontaneous humor and dramatization. For example, rather than reading a poem, he sings it, or he acts excerpts of a play. He poses questions that stimulate students' thinking and their active participation in class discussion. He acts sympathetically toward students, and listens to them carefully and with interest.

Excerpt 6: Analysis of Classroom Videotapes of Teacher B

Teacher B uses rich and fluent language, incorporating many anecdotes, metaphors, stories from her own life, and examples from the students' world of reference and from literature with which they are already familiar. She makes comparisons and distinctions to support her explanations, repeats points she thinks may be difficult and uses the rule-example-rule technique. She promotes students' intuition by presenting elements of the stories on the board to help them visualize, and at times she refers to these drawings to indicate where they are in terms of discussing the theory. She uses much spontaneous humor, mimicry, and gestures. Her most remarkable characteristic is her outstanding enthusiasm and dynamism in presentation. She provides historical and philosophical background to the topics and often challenges students' thinking by posing intriguing questions to them. She also shows care and respect for students, covers for unclever questions, maintains good rapport throughout the lesson, and answers questions in a pleasant and encouraging way.

Excerpt 7: Analysis of Classroom Videotapes of Teacher C

Teacher C, in the first lesson of the course, distributes a printed syllabus organizing the course topics into categories and subcategories. She then discusses the rationale and structure of the syllabus with the students. She links the lesson topic to the overall course framework by often saying things like: "The fourth determinant is . . . , and I remind you that we are dealing now with intellectual aptitudes," or "Don't lose sight of the whole picture—we are now in the category of emotional motives." She presents the lesson outline right at the start of the lesson, and clearly divides the lesson into topics and subtopics. She indicates that she is moving from one part of the lesson to another by saying "We have finished talking about . . . , and now we'll talk about . . ." or "Now I want to discuss. . . ." Sometimes she marks the move by writing the headings of the new topic on the board. She brings many practical examples of personal testing. She checks students' understanding from time to time. When she concludes her answer to a student's question, she asks if additional explanation is needed. She also excels in explaining, defining, and elaborating unfamiliar or complex notions and terms. She writes on the board extensively and legibly and uses well-designed transparencies. To make her lessons engaging she presents questions that challenge students' thinking, and she actively involves students in the material so that they can experience some aspects of it for themselves. For example, when she teaches the Drawing Test (a personality test), she presents slides with drawings of children and adults and assigns students to analyze the drawings and present their conclusions to class. She also assigns students to take some of the tests during the lesson and then to score themselves on a scale she provides. They then jointly discuss the test and its scoring. She creates a very pleasant and sympathetic classroom environment. She smiles a lot, jokes and laughs with the students, comes to class 5-10 minutes before the lesson starts and stays there after the lesson ends to talk with the students. She communicates to students that she cares about them and wants them to succeed. She encourages students through positive feedback, for example: "Excellent question!" or "You have raised a very important point." She enables all students to ask questions and willingly and clearly answers them.

Excerpt 8: Analysis of Classroom Videotapes of Teacher D

Teacher D shows very low use of organization strategies, except for linking the lesson to the preceding one. Regarding clarity, he brings many relevant illustrations of political violence in the country and the world to support the points he makes and exhibits superb linguistic skills: his sentences are short and simple, he uses simple familiar words, his speech is very articulate and fluent, his voice is pleasant to listen to, and he often pauses to enable students to think or to take notes. He encourages students' questions and answers them willingly and concisely. He also uses many of the strategies that make a lesson engaging: he is very enthusiastic and dynamic with gestures and facial expressions to support what he is saying, moves a lot in the classroom, and communicates high self-confidence. His voice is well modulated and animated and he changes the rate and intonation of speech to emphasize important points. He often presents analogies, philosophical or historical notes, and relevant anecdotes, based usually on political or historical events. He uses humor extensively, in connection to the material taught. He responds spontaneously and with humor to nonteaching-related, unanticipated events. His illustrations and anecdotes are taken from daily life, and he presents them in a very dramatic and lively manner, which catches students' attention. He poses thinking challenges that demand evaluation and judgment. When the time for thinking is over, the participants present and discuss their opinions. He often uses self-disclosure, describing interesting personal or professional experiences. He strongly displays all the strategies that are known to promote classroom climate: he is relaxed, flexible, talks with students in a very pleasant and polite manner, is open to criticism, and encourages students to freely express their ideas. He generally encourages students' input by praising their questions or answers, and he maintains eye contact with the students throughout the lesson.

Source: students' ratings. Table 2 presents the means and standard deviations of students' ratings of all four teachers on the questionnaire items.

A rule of thumb I formulated (Hativa, 1995) suggests to roughly interpret the mean ratings on a scale from 1 to 5 as follows: (a) below 2.9 very low; (b) 3.0–3.4 low; (c) 3.5-3.9 medium; (d) 4.0–4.4 high; and (e) 4.5-5.0 very high. Table 2 indicates that all teachers are rated either high or (mostly) very high on the first six general items. That is, all teachers are perceived by their students as extremely well prepared, inducing an exceedingly pleasant classroom climate, and being outstandingly clear, interesting, and organized.

In order to enable a comparison of students' ratings to the other two sources for teacher evaluation, we have converted the numerical mean ratings as presented in Table 2 into their corresponding letter representation using the above rule of thumb. For example, the mean ratings of the four teachers on the item "Links to the overall course framework," as appears in Table 2, were respectively 3.73, 4.45, 4.54, 3.86. These numbers are represented in Table 1 by letters: c, d, e, c (medium, high, very high, medium). Thus, the four columns on the right in Table 1 are letters representing the information from Table 2.

Integration of both sources. Table1 integrates the results of all three sources. We compare here particularly the videotapes and student ratings and we name "highly used" those strategies rated high (3, 4 or 5; c, d or e) on both—videotapes and students. By "very highly used" we refer to strategies of the upper two levels (4 or 5; d or e). For strategies rated high or very high by a single source, we do mention the source.

Classroom strategies that are highly or very highly used by all teachers. There is only a single *organization* strategy that is very highly used on the basis of a single source of evidence (videotapes)—linking the lesson to the previous lesson. Strategies highly used by all four teachers are: linking each lesson to the overall framework of the course, and di-

TABLE 2

Students' Ratings of their Instructors on the Effective-Teaching-Behaviors Questionnaire

Department:	Literature		Psychology			
Instructor: Teaching behaviors/strategies	A Mean (*SD*) n = 34 (62%)	B Mean (*SD*) n = 100 (79%)	C Mean (*SD*) n = 81 (65%)	D Mean (*SD*) n = 73 (66%)	Per item Mean (*SD*)	Range[1]
HIGH-INFERENCE ITEMS						
Overall teaching performance	4.31 (0.64)	4.62 (0.94)	4.58 (0.56)	4.66 (0.51)	4.54	0.35
Overall lesson organization	4.26 (0.83)	4.41 (0.89)	4.72 (0.57)	4.01 (0.77)	4.35	0.71
Overall lesson clarity	4.24 (0.89)	4.53 (0.93)	4.60 (0.66)	4.53 (0.58)	4.48	0.36
Overall lesson was interesting	4.18 (0.80)	4.69 (0.85)	4.04 (0.76)	4.74 (0.47)	4.41	0.70
Induction of a pleasant classroom environment	4.29 (0.80)	4.66 (0.70)	4.88 (0.33)	4.53 (0.63)	4.59	0.59
Preparedness for lesson	4.64 (0.78)	4.72 (0.83)	4.77 (0.42)	4.85 (0.36)	4.75	0.21
Intellectual challenge & promoting self-thinking	3.84 (1.05)	4.26 (0.93)	3.80 (0.82)	4.45 (0.71)	4.09	0.65
Effective use of class time	4.21 (0.84)	4.64 (0.75)	3.93 (0.87)	4.01 (0.82)	4.20	0.71
LOW-INFERENCE ITEMS						
Lesson Organization						
Links to the overall course framework	3.73 (0.94)	4.45 (0.90)	4.54 (0.61)	3.86 (0.85)	4.15	0.81
Presents lesson outline at start	2.20 (1.19)	2.08 (1.30)	4.09 (0.72)	1.96 (0.77)	2.58	2.13
Summarizes the lesson at end	3.29 (0.97)	4.03 (1.13)	3.17 (1.04)	2.33 (0.83)	3.21	1.70
Divides material into topics	3.78 (1.07)	4.02 (1.13)	4.74 (0.47)	3.81 (0.88)	4.09	0.96
Mean per instructor	3.25	3.65	4.14	2.99	3.66	1.15
Lesson Clarity						
Explains things simply	4.00 (0.89)	4.26 (0.97)	4.58 (0.56)	4.33 (0.65)	4.29	0.58
Presents questions to check if students understand	2.48 (1.12)	2.41 (1.54)	3.99 (0.82)	3.48 (0.97)	3.09	0.49
Answers well students' questions	3.82 (1.03)	3.94 (1.18)	4.64 (0.63)	4.55 (0.65)	4.24	0.82
Emphasizes important points	4.06 (0.85)	4.63 (0.63)	4.21 (0.77)	4.22 (0.73)	4.28	0.57
Writes (legibly) on board/transparencies	2.07 (1.21)	2.18 (1.35)	4.62 (0.62)	3.52 (1.03)	3.10	2.55
Provides intuitive meaning to explanations	3.39 (1.06)	4.48 (0.81)	3.84 (1.00)	4.10 (0.80)	3.95	1.09

TABLE 2 (*Continued*)

Department:	Literature		Psychology			
Instructor: Teaching behaviors/strategies	A Mean (*SD*) n = 34 (62%)	B Mean (*SD*) n = 100 (79%)	C Mean (*SD*) n = 81 (65%)	D Mean (*SD*) n = 73 (66%)	Per item Mean (*SD*)	Range[1]
Gives appropriate examples/illustrations	3.78 (1.13)	4.60 (0.72)	4.72 (0.53)	4.60 (0.62)	4.43	0.94
Links explanations to students' prior knowledge/ life experience	2.84 (1.04)	4.05 (1.15)	3.64 (0.97)	2.85 (0.84)	3.35	1.21
Speaks intelligibly	4.21 (1.15)	4.29 (1.27)	4.36 (0.92)	4.63 (0.51)	4.37	0.42
Explains unfamiliar terms	3.16 (1.37)	4.23 (0.93)	4.59 (0.64)	4.26 (0.71)	3.88	1.43
Does not make errors/mistakes	4.50 (0.89)	4.54 (1.07)	4.70 (0.51)	4.70 (0.52)	4.61	0.20
Mean per instructor	3.81	3.96	4.35	4.11	3.99	0.54
INTERESTING/ENGAGING LESSON						
Provides motivation to study	3.33 (1.29)	4.21 (0.87)	4.05 (0.75)	4.75 (0.49)	4.09	1.42
Enthusiastic presentation	3.32 (1.27)	4.87 (0.51)	4.20 (0.97)	4.63 (0.57)	4.26	1.55
Uses analogies, metaphors, anecdotes	3.23 (1.18)	4.68 (0.73)	3.90 (0.87)	4.74 (0.50)	4.14	1.51
Provides philosophical or historical background	3.19 (1.31)	4.51 (0.84)	3.80 (0.88)	4.26 (0.75)	3.94	1.32
Mean per instructor	3.27	4.57	3.99	4.60	4.11	1.33

[1] The range in ratings on a particular item is the difference between the highest and lowest ratings among the four teachers. To illustrate, on organization it is from 4.72 –4.01 = 0.71.

viding the lesson or the topic into subtopics or theses. Regarding *clarity* strategies, teachers use very highly the following strategies: simplifying explanations, giving good examples and illustrations, avoiding errors, speaking distinctly and clearly,[1] and repeating and elaborating difficult points (videotapes). Highly used clarity strategies are: encouraging students to ask questions when they do not understand (videotapes), answering students' questions well, and emphasizing important points. For making the presentation *interesting/engaging,* no strategy shows to be highly used in either one of the sources, probably because Teacher A is rated low by students on all four items, which contradicts his ratings by the judges. However, if we refer solely to the videotapes, then motivating to study the material, presenting thinking challenges, introducing variety to the lesson, and activating students during the lesson are very highly used by all teachers, and enthusiastic/dynamic presentation and providing philosophical or historical background are highly used. When considering the videotaped lessons, all teachers appear to maintain a very pleasant *classroom climate*: they strongly demonstrate care for students and their learning, behave respectfully towards them, provide encouraging feedback, and are very approachable to them. Three of the teachers use humor in class.

In sum, regarding strategies that are very highly used, only four are supported by both sources, and they all belong to the category of clarity: simplifying explanations, giving good examples and illustrations, avoiding errors in presentation, and speaking distinctly and clearly. Additional strategies that show to be very highly used on the basis of the videotaped lessons, but for which we do not have student rating evidence (because of lack of corresponding questionnaire items), are: linking the lesson to the previous one, repeating and elaborating on difficult points, presenting thinking challenges, introducing variety, and the four strategies that promote positive classroom climate and rapport with students. We see here also some major disagreements between the judges of videotapes and students' ratings respectively, which may stem from the different type of experience with the teacher of each group of evaluators.

Strategies that are in low use by at least one teacher. Of special interest are effective teaching strategies that are only little used, if at all, which suggests that these strategies are not a necessary condition for superior teaching. We refer here as "of low use" to strategies rated at the two lowest levels: either 2 or 1, or b or a, for at least one teacher. Low-use strategies supported by both sources are: outlining and summarizing the lesson, writing (legibly) on the board or on transparencies, and including intuitive meaning in explanations. Low-use strategies supported solely by the videotaped lessons, that is, for which we do not have student ratings' evidence, are: marking the move from one topic to the other, using teaching aids, presenting contributions from own professional or personal experiences, integrating material from other domains, and using humor.

Differences among teachers in use of effective classroom strategies. Tables 1 and 2 indicate that each of the four teachers achieves his or her high effectiveness in different ways, and each emerges to have his or her own strengths and weaknesses. We next examine these differences separately for the general items and for the specific items-the effective classroom teaching strategies. Because all teachers score very high on classroom climate, we will not refer to this category in the discussion.

General teaching dimensions. Table 2 shows that the smallest range size in ratings on the eight general and main dimensions is on lesson preparedness, and then on overall teaching performance and on lesson clarity. This small range indicates that there is hardly any difference in students' very high satisfaction from the instruction of the four teachers and that all four are perceived as extremely well prepared for class. Of the four main dimensions, the range size in ratings on clarity is substantially smaller than on all others, suggesting that all teachers exhibit a similar level of clarity, whereas they differ on levels of organization, interest, and intellectual challenge. Teacher A is rated high on almost all general items, except intellectual challenge (medium-high). Teacher B particularly excels on lesson interest, a pleasant classroom climate, and effective use of class time. Teacher C is rated very high on lesson organization and on pleasant classroom climate, significantly higher than all others are. She is also rated the highest on clarity, although the gap between her and the others is small. On the other hand, she is rated the lowest on lesson interest/ engagement, intellectual challenge, and effective use of class time. Teacher D is rated the highest on overall performance, lesson preparedness, making the lesson interesting, and providing intellectual challenge, but is rated the lowest on lesson organization.

Effective classroom strategies. When we use classroom videotapes, Table 1 reveals large differences among teachers in level of use of the different strategies. Rather than discussing each listed strategy separately, we refer here to the mean ratings on strategies under the main dimensions of effective teaching. Table 1 indicates that while the mean rating on strategies of lesson clarity is very similar for all teachers (range size of 4.2–3.8 = 0.4), the gap in means is relatively large on organization and interest (respectively 1.7 and 1.1). Regarding organization, Teachers A and C rate high while Teachers B and D rate low, whereas on lesson interest the situation is quite the opposite—Teachers B and D rate very high, whereas Teachers A and C rate medium.

When we use students' ratings of teachers, Table 2 also reveals a large variation among teachers in level of use of classroom strategies. The mean ratings per teacher of strategies belonging to each main dimension suggest again no substantial differences on clarity (range size: 4.35-3.81 = 0.54) and much larger differences on organization and interest (respectively 1.15 and 1.33). These means also support the high performance of all teachers on clarity, the particular excellence of Teacher C in organization, and of Teachers B and D in making the lesson interesting/engaging. To conclude, although all four teachers are rated high on the main dimensions of effective teaching—they are well organized, clear and interesting, and establish a very pleasant learning climate, they differ on the classroom strategies that they use to achieve the main dimensions. Interestingly, although intuitively predictable, teachers high on a particular main dimension of effective teaching also use effectively the classroom strategies that contribute to this dimension. These findings validate students' ratings as well as the definitions of components of the main dimensions of effective teaching. Another important conclusion is the similarity in high extent of use of clarity as a main dimension and as classroom behavior among all teachers.

When we compare the three sources for teacher evaluation (Table 1 mean ratings of main and general dimensions and of strategies, Table 2 mean ratings of strategies), we get pretty high consistency on many of the evaluations. All three sources indicate high or very high ratings of Teacher C in organization, of Teachers B and D in interesting/engaging lessons, and of all teachers on clarity. There are only few inconsistencies, particularly on organization. Teacher A's ratings in Table 1 on both the general dimensions and the particular strategies are high, whereas on the strategies in Table 2 they are low. For Teacher B these are respectively very high, low, and medium, and for Teacher D—very high, low, and low. The main difference seems to be between mean ratings on classroom strategies and ratings on the general dimensions. It is possible that the strategies selected to represent the main dimension of organization do not fully represent this notion.

Third Question: What are the Relationships Between Teachers' Beliefs and Knowledge About Effective Teaching Stràtegies, and their Classroom Use of These Strategies?

Comparing results of analyses of teachers' interviews and of the videotaped lessons (Table 1) reveals a sound number of strategies that our exemplary teachers actually used at a high level (rated as either 4 or 5), even though they did not mention them in their interviews, in spite of prompts. This is particularly the case for strategies in the organization and clarity domains in which teachers failed to recognize almost one half of the strategies that they actually used (6 out of 14 occurrences—43%, and 19 out of 37 occurrences—51%, respectively). For example, in her interview, Teacher B declared herself to be disorganized and said she did not let students ask questions in class, and that she rather asked them to submit in writing. However, her class videotapes show that she did enable students to ask questions during the lesson, that her organization was felt through many connections that she made to the general framework of the course and the previous lesson, through her spiral advancement in the material, and through her concentration on central theses, which gave a sense of continuity and coherence in the topic presentation. A comparison of results of teachers' interviews with those of the students' ratings leads to similar conclusions.

On the other hand, in some cases (two in the organization category, four in the clarity category and two in the interesting/engaging category) the teachers mention in their interviews using strategies that either one or both sources—the students and the judges of their videotaped lessons—evaluate as of very little use, if at all. For example, Teacher A thinks he summarizes the lesson at its end, whereas neither the judges nor the students rate him high on this strategy. Similarly, Teachers B and D believe that they present sufficient questions to check students' understanding, in contrast with students' ratings.

Differences Between the Two Departments.

Two teachers per department form too small a number to establish disciplinary differences. The following is a qualitative evaluation for differences of this type. In Table 1, there is a single item that clearly differentiates between the two departments—"Writing on the board or using transparencies." Both judges and students rated the two psychologists as medium and very high, respectively, on this item, whereas they rated both literature teachers very low. These ratings suggest that literature teachers may use transparencies or writing on the board substantially less than psychology faculty. The student ratings suggest a few additional differences: the psychologists seem to present a lesson outline at the beginning, whereas this is not found for literature teachers, and the psychologists present questions to check understanding and answer students' questions somewhat more than do the literature teachers. One interesting out-of-class behavior that is particular to the two literature teachers is going out with the students to see a play.

Summary and Discussion

We next summarize and discuss the findings regarding the three research questions.

First Question: What do Exemplary Teachers Believe and Know About Effective Teaching Dimensions and Strategies?

All four teachers recognize the importance of clarity, of providing motivation for learning, and of creating a pleasant classroom climate that is conducive to learning. However, lesson organization is regarded as not very important for effective teaching by two of the teachers. The four participants are familiar with a sound number of effective classroom strategies but there is a large variation in knowledge among these teachers. They show relatively low knowledge of teaching strategies that lead to lesson organization and clarity, and much higher knowledge of those that make a lesson interesting and engaging and create a positive classroom climate. In addition, the teachers are unaware of a substantial number of important strategies, including some of those they actually use in class.

We do not have in this study any data to compare our teachers' knowledge to that of novice teachers, as suggested in the summary of literature above, but we can make a limited comparison with poor teachers using three case studies (Hativa, 1998; in print). These studies indicate that the low-rated teachers involved were familiar with only few effective teaching strategies. Thus, maybe one way in which the pedagogical knowledge of good and poor teachers differs is in the number of effective classroom strategies with which that they are familiar.

Second Question: What Strategies Do the Exemplary Teachers Actually Use in Class?

All our teachers show about the same high (but not very high) level of clarity and very high level of pleasant classroom climate. However, two teachers excel in course and lesson organization, while the level of interest and engagement in their lesson is in the medium-high range. By contrast, the other two excel in making their lesson interesting and engaging but are low or even very low on use of strategies for lesson organization. Thus, none of our four exemplary teachers excels in all main dimensions of effective teaching. This finding is supported by two other studies in which different research methods were used.

Lowman (1996) suggests a model of effective teaching that consists of only two main dimensions: (a) good presentation skills: offering presentations in clearly organized and interesting ways, and (b) interpersonal rapport: relating to students in ways that communicate positive regard and motivate them to work hard to meet academic challenges. Thus, Lowman's first dimension combines our three dimensions of organization, clarity, and interest/engagement, whereas his second dimension corresponds to our dimension of classroom environment. Lowman found that each of the many exemplary teachers he had studied was at least moderately successful at the skills needed for one of the two dimensions and exceptionally skilled in the other, with only very few teachers being exceptionally skilled in both dimensions. Our study, although using entirely different research methods, has arrived at a similar conclusion, only we refer to four main dimensions rather than the two dimensions mentioned by Lowman.

Young and Shaw (1999) also support these findings. They asked a large number of students to rate a teacher of their choice, from whom they had taken a college or university course in their recent past, on a 25-item instrument. On the basis of various data analyses, including clustering procedures, they produced six main dimensions of effective teaching: value of the course, motivating students, comfortable learning atmosphere, course organization, effective communication, and concern for student learning. If we combine their third and last dimensions, then all their dimensions except of the first one are parallel to the four main dimensions in our study: interest, clarity, organization, and a positive classroom climate. Their data analysis also produced five distinct profiles of effective teachers. One profile represented teachers who were high on all of their main dimensions. All other profiles included a single dimension on which the teachers were rated substantially lower than they were in the other dimensions. Thus, second-profile teachers were relatively low on communication, third-profile on organization, fourth-profile on motivation, and fifth-profile on comfortable classroom atmosphere.

All these studies provide validity to the notion that although effective teaching is considered to consist of high organization, lesson interest and clarity, and of a positive classroom climate, even exemplary teachers may not perform on the highest level in all these categories, and they may be medium or even low on some of the main dimensions of effective teaching, provided that they are high on all others.

Our teachers were also found to use different classroom strategies. Only a few strategies were highly used by all teachers. Several of these were already mentioned in the studies reviewed in the introduction, for instance, the use of good/proper examples and illustrations, simplifying explanations, speaking intelligibly, and avoiding making errors.

This study shows that a teacher can excel even without using some of the strategies regarded as very important for effective teaching, such as using teaching aids, writing on the board, outlining and summarizing the lesson, or using humor. Thus, we may perceive use of these strategies as a sufficient but not necessary condition for exemplary teaching.

We may conclude that there is no single way, no particular comprehensive set of classroom strategies that is necessary for becoming an exemplary teacher. Each teacher has his or her own profile of strategies that he or she applies in class and that contributes to teaching effectiveness. Indeed, there is a small set of strategies that seem to be of high importance for effective teaching. However, beyond this small set, each teacher achieves his/her excellence in a different way, using a different combination of classroom strategies. An instructor can perform medium or even low on many classroom strategies of effective teaching and still be perceived by students as an excellent teacher, provided that he or she uses some of the other strategies in a very effective way.

Third Question: What are the Relationships Between Exemplary Teachers' Beliefs and Knowledge About Effective Teaching Strategies, and Their Actual Use of These Strategies?

Our teachers apply in practice many of the effective teaching strategies they are familiar with. For example, Teacher A, as his interviews show, believes that organization and a good student-teacher relationship strongly increase teaching effectiveness, and so he implements these ideas in his teaching. Hence his lectures are well organized, clear, and interesting, and he maintains good rapport with students. Teacher B believes that preparation before class, being clear and interesting, and having a good rapport with the students are important factors in effective teaching. Indeed, in her teaching she focused on clarity and on presenting the lesson in an interesting manner, and she maintained good interaction and rapport with her students.

In addition to those strategies that our teachers knew and did use in class, there was a good number (about 50%) of strategies which they used a great deal, but which they were not aware of using. They did not even mention these strategies upon being prompted during their interviews. We may assume that our teachers used some effective strategies by intuition, with probably only an implicit knowledge of them. Still as a third case, several of the effective strategies our teachers mentioned in their interviews did not show in their videotaped classes and were not mentioned by their students either, suggesting that even exemplary teachers do not apply in class all effective teaching strategies with which they are familiar. These findings support Thompson's (1992) findings of inconsistent relationships between university teachers' knowledge and beliefs regarding teaching and their observed practice. The literature presented above explains this discrepancy by the many factors influencing teacher work, such as the social context in which teaching takes place, including the values, beliefs, and expectations of peers, academic administrators, and students; or the lack of intentional linking of knowledge and experience to action through reflection, practice, and feedback over time. Another explanation, based on studies of poor teachers, (Hativa, 1998; in print) related this discrepancy to the professors' damaging thinking and beliefs about teaching, learning, and students, and to some of their personal characteristics which badly affected good instruction.

Conclusions and Implications

This study supports the intuitive belief of education professionals that different exemplary teachers achieve effectiveness in very different ways. Polya, the famous mathematician and mathematics educator, expressed this belief many years ago: "I can give you no rules [for effective teaching], for there are as many good ways of teaching as there are good teachers" (Polya, 1957, p. 37).

Our study suggests that to be an exemplary teacher, one does not need to excel in all four main dimensions of effective teaching (organization, clarity, etc.). It is sufficient to excel in some of these dimensions and to function quite well on the others; one can even function poorly on one of these. Clarity and a positive classroom climate emerge as the two out of the four main dimensions of effective teaching in which all our teachers showed at least a high level of performance, and these findings are triangulated by several sources.

The study indicates that our exemplary teachers' beliefs concerning, and knowledge of, effective teaching strategies, although developed to a high degree, are far from being fully developed. It also shows that there is a good, but far from perfect, fit between these teachers' beliefs and knowledge concerning effective strategies and their classroom practice. On the basis of previous literature we suggest that there is much less of such fit for less good teachers, and that the reasons for this lower correspondence should be further investigated.

The results of this study may serve instructional consultants who work with faculty members on improving their teaching and can help more directly those faculty members who are working on improvement. The main idea is that an essential aspect of preparing faculty members for their teaching role is to increase their knowledge of a wide variety of teaching strategies and to help them understand how each of these strategies contributes to the main dimensions of effective teaching: organization, clarity, and so on. Having gained this basis, each teacher should select those strategies that fit his or her personality, skills, thinking and beliefs, subject matter, students in class, and other factors of the particular teaching context. When teachers reflect on their instruction, they should consider whether they have actually been using, in class, all the strategies they planned to use and how successful they were in using them and in achieving organization, clarity, and so on. We believe that even the highest rated teachers may improve by adding effective strategies to their repertoire of classroom behavior and that they should examine their actual implementation. However, for those teachers perceived by their students as less than exemplary, and particularly for poor teachers, we suggest that other people than they themselves—such as colleagues or instructional consultants—supervise and follow up the implementation in class of the strategies these teachers have selected for their use.

Limitations of the Study and Suggestions for Further Study

The small number of teachers involved is the main limitation, although each case study took a very large amount of time and work, thus enabling us to investigate in depth the topics involved. Our study should therefore be regarded as only an initial step in answering the questions posed. Additional case studies are needed to establish our results, using exemplary teachers as well as less good teachers from different universities and colleges and from different departments. Further, larger-scale studies should establish whether clarity and a positive classroom climate are prerequisite dimensions of good teaching and whether the classroom strategies that we identified as highly used by all our teachers form a necessary condition for effective teaching.

Notes

[1] This strategy is included in the "very highly used" category in spite of the single c, because six of the eight ratings are at the top–5 or e.

References

Andrews, J., Garrison, D., & Magnusson, K. (1996). The teaching and learning transaction in higher education: A study of promoting excellent teaching. *Teaching in Higher Education, 1*(1), 81–103.

Berman, J., & Skeff, K. M. (1988). Developing the motivation for improving university teaching. *Innovative Higher Education, 12*(1), 114–125.

Blackburn, R. T., Boberg, Q., O'Connell, C., & Pellino, G. (1980). *Project for faculty development program evaluation: Final report.* Ann Arbor, MI: Center for the Study of Higher Education, University of Michigan.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26,* 473–498.

Centra, J. A. (1996). Identifying exemplary teachers: Evidence from colleagues, administrators, and alumni. *Honoring exemplary teaching.* New Directions for Teaching and Learning, no. 65, pp. 51–56.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255-296). New York: Macmillan.

Cross, K. P. (1977). Not can, but *will* college teaching be improved? In J. A. Centra (Ed.), *Renewing and evaluating teaching.* New Directions for Higher Education, no. 17, pp. 1–15. San Francisco: Jossey-Bass.

Dunkin, M. J. (1995). Concepts of teaching and teaching excellence in higher education. *Higher Education Research and Development, 14*(1), 21–33.

Dunkin, M. J., & Precians, R. P. (1992). Award-winning university teachers' concepts of teaching. *Higher Education, 24,* 483–502.

Feldman, K. A. (1989a). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30,* 583–645.

Feldman, K. A. (1989b). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education, 30,* 137-194.

Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147-164). New York: Macmillan.

Gudmundsdottir, S., & Shulman, L. (1987). Pedagogical content knowledge: Teachers' ways of knowing. *Scandinavian Journal of Educational Research, 31*(2), 59–70.

Hativa, N. (1995). The department-wide approach to improving faculty instruction in higher education: A qualitative evaluation. *Research in Higher Education, 36,* 377-413.

Hativa, N. (1997). *Teaching in a research university: Professors' conceptions, practices, and disciplinary differences.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Hativa, N. (1998). Lack of clarity in university teaching: A case study. *Higher Education, 36,* 353–381.

Hativa, N. (1999). Towards a conceptual framework of dimensions of effective instruction: The role of high- intermediate- and low-inference teaching behaviors. *Instructional Evaluation and Faculty Development, 18,* 3–10.

Hativa, N. (2000a). Clarity in teaching in higher education: Dimensions and classroom strategies. In D. Lieberman (Ed.), *To improve the academy* (Vol. 19). Bolton, MA: Anker.

Hativa, N. (2000b). *Teaching for effective learning in higher education.* Dordrecht, Holland: Kluwer Academic Publishers.

Hativa, N. (2000c, April). *The tension between thinking and beliefs of professors and students regarding instruction.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Hativa, N. (in print). Becoming a better teacher: A case of changing the pedagogical knowledge and beliefs of law professors. In N. Hativa & P. Goodyear (Eds.), *Teacher thinking, beliefs and knowledge in higher education.* Dordrecht, Holland: Kluwer Academic Press.

Hilgemann, V., & Blodget, J. (1991). *Profile of an excellent teacher: Summary of research* (ERIC Document Reproduction Service No. ED 353017).

Horan, M. (1991). *Attributes of exemplary community college teachers: A review of the literature* (ERIC Document Reproduction Service No. ED346900).

Kelly, N., & Kelly, B. (1982, November). *Backgrounds, teaching styles of award winning professors.* Paper presented at the annual meeting of the Rocky Mountain Educational Research Association, Albuquerque, NM.

Lowman, J. (1996). Characteristics of exemplary teachers. New Directions for Teaching and Learning, no. 65, pp. 33–40.

McAlpine, L., & Weston, C. (in print). Reflection: Issues related to improving professors' teaching and students' learning. In N. Hativa & P. Goodyear (Eds.), *Teacher thinking, beliefs and knowledge in higher education.* Dordrecht, Holland: Kluwer Academic Press.

Murray, H. G. (1997). Effective teaching behaviors in the college classroom. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 171–203). New York: Agathon.

Ovando, M. N. (1989). *An effective faculty development program: It can be done.* Paper presented at the annual conference of the National Council of States on Inservice Education, San Antonio, Texas.

Peterson, P. L. (1988). Teachers' and students' cognitional knowledge for classroom teaching and learning. *Educational Researcher, 17*(5), 5-14.

Polya, G. (1957). *How to solve it? A new aspect of mathematical method* (2nd ed.). Princeton, NJ: Princeton University Press.

Ropo, E. (1987, April). *Teachers' conceptions of teaching and teaching behavior: Some differences between expert and novice teachers.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Stader, E., Colyer, T., & Berliner, D. (1990, April). *Expert and novice teachers' ability to judge student understanding.* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-130). New York: Macmillan.

Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education, 42,* 292–305.

Wilson, S. M., Shulman, L. S., & Richert, A. E. (1987). '150 different ways' of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp. 104–124). London: Cassell Educational Limited.

Young, S., & Shaw, D. G. (1999). Profiles of effective college and university teachers. *Journal of Higher Education, 70,* 670–686.

JE John Robst

Cost Efficiency in Public Higher Education Institutions

"South Carolinians have come to doubt that public colleges are using their tax dollars wisely, . . . and this has made lawmakers reluctant to spend more on higher education" (Austin Gilbert of South Carolina's Commission on Higher Education, quoted in Schmidt, 1997, p. A26).

Introduction

South Carolina is one of many states to question whether higher education institutions operate efficiently. Government officials suggest that using performance measures and limiting state appropriations for higher education will control higher education costs and force institutions to provide an education more efficiently. For example, the 1999 Almanac issue of *The Chronicle of Higher Education* reported that 22 states provide appropriations to individual public colleges partly based on their meeting specific performance measures. Although state appropriations for higher education have generally increased in recent years due to a strong economy, it is a matter of time before the economy slows. In some cases, even a strong economy is not enough to increase state appropriations. For example, Governor Pataki's 1999 budget proposal sought reductions in state appropriations for the State University of New York and City University of New York systems. Appropriation

---

An earlier version of this article was presented at the 1997 Eastern Economic Association meetings. I would like to thank conference participants, Jack Keil, Robert Roseman, and two anonymous referees for insightful comments. The opinions in this article are the author's and do not represent those of the Health Care Financing Administration.

John Robst is an economist at the Centers for Medicare and Medicaid Services, Baltimore, MD.

The Journal of Higher Education, Vol. 72, No. 6 (November/December 2001)
Copyright © 2001 by The Ohio State University

reductions were sought to encourage what Governor Pataki called "continued efforts to improve cost-efficiency and productivity" (*Chronicle of Higher Education's* Daily Report, January 28, 1999).

Why some universities are more efficient at providing public higher education to students and the relationship between efficiency and an institution's revenue structure are not adequately addressed in the literature. This article attempts to partly fill this gap with a three-step analysis. First, the changed structure of public university revenues between 1991 and 1995 is documented. Reductions (or limited growth) in state appropriations combined with increases in tuition revenue resulted in universities receiving a smaller proportion of revenues from state appropriations and a larger proportion from tuition.[1] The second step analyzes how this changed revenue structure influenced cost efficiency defined as the difference between actual costs and an estimated minimum cost. Attention is focused on whether public sector universities that received a larger proportion of funds from state appropriations were relatively more or less efficient than schools that received a larger share of funds from other sources such as tuition.[2] The third step examines how universities' cost efficiency and revenue structure changed between 1991 and 1995. In particular, this paper determines whether schools with a greater decline in the share of revenues from state appropriations improved efficiency relative to schools with a smaller decline in the share from state appropriations.

Literature Review

Before getting to these questions, we discuss some of the literature examining higher education costs (e.g., Adams, Hankins, & Schroeder, 1978; Brinkman, 1981; Brinkman & Leslie, 1986; Brovender, 1974; Hoenack, Weiler, Goodman, & Pierro, 1986; Koshal & Koshal, 1995; Tierney, 1980; Verry & Layard, 1975). Early studies concentrate on a single output in higher education, namely undergraduate students measured by full-time equivalent students, number of credit hours, or number of students. This is problematic, for colleges and universities have various different outputs. James (1978) noted that focusing only on undergraduate students and ignoring other higher education outputs, such as graduate education and research, overstates the cost of providing an undergraduate education and creates biases across types of institutions.

Recent studies overcome this problem by considering universities to be multi-product firms producing undergraduate education, graduate education, and research. Cohn, Rhine, and Santos (1989) use Higher Education General Information Survey data on more than 1,000 institutions to estimate cost functions for public and private schools. De Groot, McMahon, and Volkwein (1991) extend this research by focusing on a subset of doctoral granting universities, using a different measure of research output, and testing for the impact of state regulation on costs. Dundar and Lewis (1995) examine costs at the department level for 18 public research universities. In general, each study finds that total costs increase as enrollment and research output increases, and that interactions between the various outputs are statistically important in determining institutional or departmental costs. In addition, economies of scale are generally found in the higher education industry.

This article takes a different approach than previous studies by estimating a frontier cost function.[3] The frontier cost function is a frequently utilized tool for looking at production costs in manufacturing plants and industries. Regression analysis typically focuses on the average producer and can be used to predict the typical producer's cost for a given output. Instead of looking at typical costs, frontier estimation predicts the *minimum* cost for producing a given amount of output, which in the case of universities is measured by students and research. The difference between the predicted minimum cost and actual cost is considered to be "excess" cost. An institution's excess costs are compared to the share of revenues from state appropriations, using appropriate controls, to determine whether shifting from state appropriations to tuition revenue affected efficiency.

This analysis uses cross-sectional data to examine differences between institutions, and time series data to examine changes within institutions. Cross-sectional analysis allows us to determine whether institutions with a smaller share of revenues from state appropriations are more or less efficient than institutions with a larger state share of revenue. However, cross-sectional analysis has two important drawbacks. First, it is difficult to infer causation[4] and second, omitted variables can bias the estimated coefficients. For example, institutions may have costs above the minimum in order to provide a higher quality education. Without adequate controls for quality, which the data do not contain, a cross-sectional approach may attribute expenditures for higher quality to inefficiency. If we assume quality changes slowly, a time series analysis can implicitly account for expenditures above the minimum that create higher institutional quality. Thus, while *levels* of inefficiency may be partly due to quality, *changes* in inefficiency over a short period of time are unlikely to be severely affected by quality considerations.

Although the relationship between a university's revenue structure and efficiency is not addressed in previous research, a number of studies provide intuitive reasons to expect a relationship between state appropriations and efficiency. For example, increasing state appropriations may lead to greater university expenditures and unless accompanied by an output or quality increase will likely lead to greater inefficiency.[5] McPherson and Shapiro (1994) find an increase in FTE state appropriations raises FTE instructional expenditures when controlling for tuition revenue and enrollment growth. In addition, Volkwein (1987) using an appropriate multivariate model shows that higher FTE state appropriations increase administrative expenditures and administrative salaries, while reducing the percentage of revenue from nonstate sources. Though not explicitly considering higher education, Gamkhar and Oates (1996) find evidence that reductions in federal non-matching grants to state and local governments lead to reduced state and local expenditures. This finding suggests that state appropriation reductions might force public higher education institutions to become more efficient. Of course, an alternative interpretation is that schools may be forced to offer a lower quality education in order to reduce expenditures.

In addition to examining how state appropriations relate to administrative expenditures and nonstate revenues, Volkwein (1987) also discusses how government regulation might affect the cost structure of universities. He finds that government regulation is positively correlated with the share of revenues from state appropriations, suggesting that institutions have greater flexibility when the state share of revenues is lower. However, Volkwein (1986a, 1986b) does not find a relationship between government regulation and FTE expenditures, nor does he find that autonomy influenced administrative costs, faculty quality, student quality, or government grants. These results indicate that a relationship between efficiency and the campus autonomy associated with a lower state share, though theoretically plausible, is unlikely.

The remainder of the article is organized as follows. The next section discusses the data used in the analysis, followed by a discussion of the methodology, results, and some concluding remarks.

Data

The data used in this article are derived from the 1991 through 1995 Integrated Postsecondary Education Data System (IPEDS). The sample is limited to four-year public institutions in the Carnegie classifications of Research, Doctoral, Masters, and Baccalaureate. Data on an institution must be available in all five years to be in the sample, because a balanced panel (the same number of observations for each institution) is necessary for estimation purposes. The time frame examined ends in 1995, because states increased state appropriations substantially in subsequent years. An interesting question for future research is whether cost efficiency has increased or decreased due to the substantial growth in state appropriations during recent years. The final sample contains 440 four-year public institutions for five years for a total of 2,200 observations. Variable means are reported in Table 1.[6] The average institution had general and educational expenditures in excess of $129 million (in 1991$) per academic year, of which $55 million came from state appropriations.[7] The share of revenue (with revenue defined as state appropriations plus tuition revenue) from state appropriations averaged 0.63 over the five years.[8]

One goal of this article is to document the shifting emphasis from state appropriations to tuition revenue. Table 2 examines institutions' changing revenue structure and helps establish the declining importance of state appropriations as a funding source. Among the 440 institutions, inflation adjusted tuition revenues increased for 426 between 1991 and 1995, while only 207 had higher inflation adjusted state appropriations. In addition, tuition revenue growth was higher in institutions with falling state appropriations, averaging $3.5 million for schools with increasing state appropriations and $6.0 million for schools with falling

TABLE 1
Variable Means

	Mean
Education and general expenditures (in thousands)	$129,063
Carnegie classification	
Research institution	0.19
Doctoral institution	0.15
Masters institution	0.53
Bachelors institution	0.13
Outputs	
Undergraduate students	11,184
Graduate students	3,064
Research expenditures (in thousands)	$20,864
Input price	
Compensation (in thousands)	$84,391
Revenues	
Tuition revenue (in thousands)	$30,237
State appropriations (in thousands)	$55,006
State share of revenue[a]	0.63
Number of institutions	440
Number of observations	2200

NOTE: All dollar values are deflated to 1991 dollars using the Higher Education Price Index.
[a] State share of revenue is measured as state appropriations divided by revenue (defined as state appropriations plus tuition revenue).

appropriations. The share of revenue from state appropriations decreased for 405 of the 440 schools. It is important to note that although many states reduced real higher education appropriations, this did not necessarily result in reduced revenues. The combination of appropriations and tuition revenue increased for the vast majority of institutions (372 of 440). Thus, consistent with other studies' findings (e.g., Healy & Schmidt, 1997), higher education costs were partially shifted from the general population to students and their parents.

Analytic Methodology

Two techniques are commonly used to examine firm or industry efficiency. Stochastic frontier estimation, first developed by Aigner, Lovell, and Schmidt (1977), is well established in the literature as an econometric method to examine costs. Data envelopment analysis (DEA) is a mathematical programming approach that compares institutions to the least cost institution. There are strengths and weaknesses to both approaches. For example, stochastic frontier estimation requires assumptions about the distributional form of the error terms, whereas DEA analysis does not. However, once the distributional form of the error terms is chosen, stochastic frontier estimation easily handles random noise caused, for example, by measurement error. Random noise can cause significant problems with DEA analysis, influencing the shape of the frontier and an institution's estimated efficiency. Given the fairly large number of institutions in the sample, the probability of random variability across institutions is too great to ignore. Thus, this article estimates stochastic frontier cost functions. Those interested in a more thorough discussion of stochastic frontier estimation, and a comparison of stochastic frontier estimation and DEA should refer to Coelli, Rao, and Battese (1997).

TABLE 2

Changes in Inflation Adjusted Tuition Revenue and State Appropriations (1991 compared to 1995)

	TUITION REVENUE		
	Increased	Decreased	TOTAL
STATE APPROPRIATIONS			
Increased	202	5	207
Decreased	224	9	233
TOTAL	426	14	

NOTE: State appropriations and tuition revenues are deflated to 1991 dollars using the Higher Education Price Index.

Begin by considering the following equation:

$$C^p_{it} = X_{u}\beta + \nu_{it} \leq C^o_{it}, \tag{1}$$

where the dependent variable represents university *i*'s minimum potential cost, X is a vector of variables measuring output quantities and input prices, v is the random disturbance distributed as $N(0,\sigma^2_v)$, and t denotes time. The minimum cost (C^p) is less than or equal to the actual general and educational expenditures (C^o) on public higher education, with the institutions operating at the efficient level having costs equal to the minimum predicted cost. However, the vast majority of universities operate above the frontier.

By reconsidering equation (1) we can see:

$$C^o_{it} = C^p_{it} + u_{it}, \tag{2}$$

where $u_{it} \geq 0$. If the university achieves its minimum cost $u_{it} = 0$. Thus, u_{it} is considered an institution specific time-varying estimate of "excess" cost at time t. Combining equations (1) and (2) leads to:

$$C^o_{it} = X_{it}\beta + \varepsilon_{it}, \tag{3}$$

where $\varepsilon_{it} = v_{it} + u_{it}$ and u is distributed half-normal, $u_{it} \geq 0$.[9] The independent variables (X) include undergraduate student enrollments, graduate student enrollments, research expenditures, faculty average compensation (salaries plus total fringe benefits), and categorical variables denoting Carnegie classification. The specification includes quadratics for each continuous variable and six interaction terms (undergraduate·graduate, undergraduate·research, undergraduate·compensation, graduate·research, graduate·compensation, research·compensation) to account for possible nonlinearities. Cohn, Rhine, and Santos (1989) use a very similar specification to estimate an OLS cost function.

Several changes to the specification are desirable, but cannot be incorporated with these data. It would be helpful to have a better measure of research output, or measures of additional outputs and institution quality.[10] Higher education costs may depend on the institution's quality, with universities offering a higher quality education incurring higher costs. Koshal and Koshal (1995) use SAT scores as a measure of higher education quality. However, SAT scores measure the quality of incoming students or elementary and secondary schooling, not necessarily higher education quality in the sense of how much a college enables students to learn. Public service outputs and capital inputs to the production process are also difficult to measure and not included in the analysis. Previous research also lacks these data.

Figure 1 provides a simple graph for noneconometricians to understand the concept of frontier analysis. Figure 1 graphs the relationship between total expenditures and "output" (assuming a single measure of output could be derived). As noted above, the OLS regression estimates the relationship between output and average costs, while the frontier predicts the relationship between output and minimum costs. Thus, the frontier predicts a lower cost at each level of output relative to OLS. Institution A on the graph has costs below the frontier. This is possible because frontier estimation allows for random variation with $v_{it} < 0$ in this case. Institution B's costs are below the average cost implying the average institution with the same outputs would have higher costs, but institution B's costs are above the frontier or minimum expected cost. If we assume for simplicity that $v_{it} = 0$, then the institution has excess costs with u_{it} equal to the difference between point B and the frontier. Lastly, institution C has costs above both average costs and predicted minimum costs. Again, if $v_{it} = 0$, then the institution has excess costs with u_{it} equal to the difference between point C and the frontier.

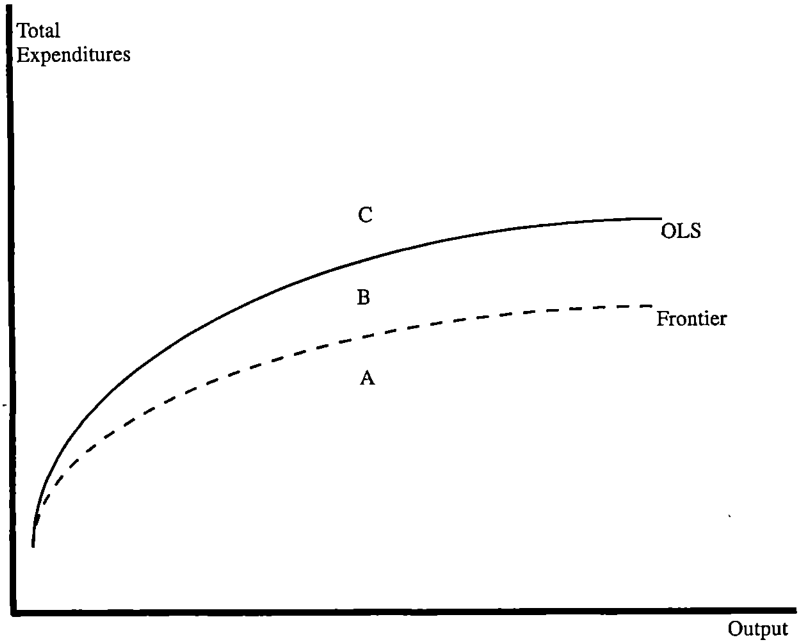


FIG. 1. An Illustration of the Cost Frontier

The primary goal of this article is to examine factors that may affect u_{it}. Prior studies that examine the causes of inefficiency in other industries use a two-step process. The first stage involves estimation of the frontier, followed by a second stage with estimated inefficiency, u_{it}, regressed on a vector of characteristics that determine inefficiency. This two-stage approach may be problematic because the inefficiency effects are not independent in the two stages (Coelli, 1996). A simultaneous equations approach developed by Battese and Coelli (1993, 1995) and Coelli (1996) uses one-step maximum likelihood estimation of the frontier and inefficiency models. In essence, the standard likelihood function used to estimate cost frontiers is expanded to model inefficiency, u_{it}, as a function of variables:

$$U_{it} = \zeta_{it}\delta + \alpha_{it} \tag{4}$$

For the purposes of our test, ζ_{it} includes the proportion of revenues from state appropriations and α_{it} is a random variable defined by the truncation of the normal distribution with zero mean and variance σ^2.[11]

Results

The results from both OLS and maximum likelihood estimation are in Table 3. Comparison of the OLS and frontier results provides a simple check that the maximum likelihood frontier coefficients are reasonable. Three maximum likelihood regressions are estimated, each with a different inefficiency specification (described below). While the variables used to estimate the frontier are the same in each regression, the coefficients vary depending on the specification of the inefficiency regression. At this point, it is useful to determine if the frontier is statistically different from OLS estimation. The gamma coefficients reported in Table 3 test the null hypothesis that σ_u^2 equals zero. If σ_u^2 equals zero, costs could be examined using OLS regressions. The null hypothesis is easily rejected in each specification, implying that inefficiency exists in the provision of higher education. Similar to other higher education cost studies, total educational and general expenditures are positively related to undergraduate and graduate student enrollments, research expenditures, and faculty and staff compensation.[12]

The first inefficiency specification examines inefficiency as a function of the share of revenues from state appropriations (state share) controlling for total university revenues (state appropriations plus tuition revenue). The university revenue coefficient is positive and significant, suggesting that larger schools have greater excess costs than smaller

TABLE 3
OLS and Maximum Likelihood Results—Panel Data

		Stochastic Frontier Coefficients		
	OLS	Specification 1	Specification 2	Specification 3
Intercept	−73055**	−88670**	−88673**	−88673**
	(3.02)	(87138)	(88502)	(88089)
Year	820.2**	921.0**	901.5**	900.9**
	(3.15)	(146.5)	(155.2)	(153.4)
Research institution	4269.4	4318.2**	4269.6**	4269.7**
	(1.56)	(1311)	(4268)	(2182)
Doctoral institution	1515.5	1627.9**	1515.4**	1515.3**
	(0.84)	(223.1)	(1515)	(1161)
Masters institution	92.08	328.6**	92.65**	92.80**
	(0.08)	(21.6)	(92.6)	(47.5)
Undergraduate enrollments	−0.3110	−0.2367	0.9132**	0.9276**
	(1.45)	(1.59)	(6.63)	(6.76)
Undergrad squared	0.00002**	−1.49E-06	−4.368E-05**	−4.418E-05**
	(2.05)	(0.27)	(7.05)	(7.09)
Graduate enrollments	0.8579*	1.588**	-0.6549*	−0.7187*
	(1.87)	(6.57)	(1.86)	(1.81)
Graduate squared	−0.00009**	−0.00009**	−0.00009**	−0.00009**
	(3.58)	(5.94)	(3.38)	(3.19)
Research expenditures	0.3252**	0.4764**	0.2008**	0.1936**
	(5.02)	(7.56)	(3.36)	(3.23)
Research squared	1.315E-06**	−2.184E-06**	2.254E-07	1.319E-07
	(2.31)	(3.99)	(0.32)	(0.19)
Compensation	1.391**	1.186**	1.219**	1.222**
	(42.8)	(37.9)	(47.4)	(47.0)
Compensation squared	1.238E-08	−8.348E-07**	−5.013E-07**	−4.491E-07**
	(0.08)	(6.22)	(2.75)	(2.99)
Under·grad	0.00008**	−2.31E-05	0.0001**	0.00014**
	(2.30)	(1.34)	(5.43)	(5.14)
Under·research	−1.455E-05**	−4.210E-05**	−1.102E-05**	1.080E-05**
	(2.93)	(8.92)	(2.38)	(2.28)
Grad·research	0.00008**	0.00010**	0.00004**	0.00004**
	(5.72)	(8.04)	(3.34)	(3.26)
Comp·under	1.116E-06	0.00002**	0.00001**	0.00001**
	(0.45)	(8.04)	(6.77)	(6.61)
Comp·grad	−0.00002**	−0.00004**	−0.00002**	−0.00002**
	(3.02)	(7.36)	(3.78)	(3.77)
Comp·research	−8.130E-07	2.824E-06**	−3.843E-07	−2.780E-07
	(1.48)	(5.21)	(0.59)	(0.44)
		The Determinants of Inefficiency		
Intercept		−4630.0**	−0.8163	−1.491
		(15.6)	(0.81)	(0.21)
Research institution				0.0848
				(0.07)
Doctoral institution				0.0254
				(0.02)
Masters institution				−1.100
				(0.21)

(Cont'd)

TABLE 3 (*Continued*)

	OLS	Stochastic Frontier Coefficients		
		Specification 1	Specification 2	Specification 3
Undergraduate enrollments			−6.000**	-6.052**
			(42.9)	(39.6)
Graduate enrollments			4.525**	4.685**
			(6.13)	(5.79)
Research expenditures			0.5882**	0.5939**
			(8.63)	(8.80)
State+tuition revenue		0.1837**	0.3475**	0.3444**
		(30.8)	(12.3)	(12.5)
State share of revenue		−2915.8**	−0.5402	−0.9870
		(15.5)	(0.54)	(0.21)
Sigma-squared		5.322E+08**	5.322E+08**	5.322E+08**
		(5.32E+08)	(5.32E+08)	(5.32E+08)
Gamma		0.9845**	0.9711**	0.9708**
		(628.0)	(359.6)	(349.7)
Number of institutions	440	Number of observations	2200	

NOTE: *t*-statistics are in parentheses.
$^{**}p < 0.05$. $^{*}p < 0.10$.

schools.[13] The state share coefficient is negative and significant, implying an inverse relationship between the state share and inefficiency. Of course, revenue includes only tuition revenue and state appropriations and may not be an adequate control for institution size. The second specification adds controls for each of the three outputs; undergraduate students, graduate students, and research expenditures. Inefficiency falls with the number of undergraduates, but increases with the number of graduate students and research expenditures. A positive relationship still exists between total revenues and inefficiency, but the relationship between the state share and inefficiency is no longer statistically significant. To further distinguish between research institutions and teaching institutions, the third specification adds controls for Carnegie classification. No significant efficiency difference exists between Carnegie classifications, with the output and revenue coefficients remaining very similar.

The results suggest that institutions with a smaller state share of revenues are not more efficient than institutions with a larger state share. Perhaps of equal interest is the relationship between university revenues and efficiency. Unfortunately, there are a number of possible interpretations for the revenue coefficients. First, as suggested above, total revenue may serve as an additional control for institutional size and other unmeasured variables. Institutions with the same number of students and research output are not necessarily peer institutions. Substantial variation may exist in fields taught, price indices due to location, and other factors that may result in differing revenues and expenditures. Second, given the lack of quality controls, it is possible that the excess cost estimates are positively correlated with institution quality. Thus, an alternative interpretation, which is the conventional wisdom, is that greater revenues improve the quality of higher education institutions. Evidence of such a relationship is provided by Volkwein (1986a), who finds a positive relationship between state appropriations and institution quality after controlling for several factors including campus autonomy, FTE students, and cost of living.

Another potential issue is the assumption that the cost function's coefficients are time invariant. The inefficiency estimates may be biased if the cost function coefficients change during the time period. Thus, the frontier and inefficiency specifications are estimated for each year with the state share coefficients from the inefficiency specification reported in Table 4. Again, there is no evidence that institutions with a lower state share of revenues are more efficient than institutions with a higher state share of revenues. Some evidence exists that the state share effect became more important over the five years. Controlling for revenue and outputs, the state share coefficient is insignificant in 1991 and 1992, but is negative, statistically significant, and increasing in magnitude between 1993 and 1995. Thus, institutions with a greater share of revenues from state appropriations were less inefficient (or more efficient), with the relationship becoming stronger over time. This result might imply that those schools with declining state shares during this time frame did not have a corresponding increase in efficiency. This hypothesis will be directly tested later in the article.

In addition to time invariant coefficients, the results in Table 3 also assume constant coefficients across Carnegie classifications. The inclusion of categorical variables for Carnegie classification allows the cost frontier to shift but requires a constant slope that may cause inefficiency to be mismeasured. To examine this issue, the sample is segmented by Carnegie classification to test whether the relationship between state share and inefficiency differs by Carnegie classification. Two specifications are estimated, one without output controls and a second including output measures (undergraduate and graduate students, and research expenditures). These results are presented in Table 5, with no significant relationship between state share and inefficiency for Research or Bachelors institutions. A marginally significant negative relationship between the state share and inefficiency exists for Doctoral institutions when controlling for outputs. Masters institutions with a larger state share are more ef-

TABLE 4

Stochastic Frontier Estimation Results by Year

	1991		1992		1993		1994		1995	
State share	−0.8968	0.0011	−1660.5**	−0.0554	−0.0075	−717.8**	-0.0086	−1018.4**	−0.0023	−1160.1**
	(0.008)	(0.001)	(4.23)	(0.003)	(0.07)	(4.58)	(0.008)	(3.73)	(0.002)	(3.96)
Other Controls	no	yes	no	yes	no	yes	no	yes	no	yes
No. of institutions	440		440		440		440		440	

NOTES: Two sets of simultaneous equations are estimated for each year with only the inefficiency specification results reported. The first specification does not control for other factors, while the second specification includes a vector of control variables. The other controls are undergraduate and graduate enrollments, research expenditures, and revenues. Both specifications include an intercept. *t*-statistics are in parentheses.

** $p < 0.05$. * $p < 0.10$.

ficient in the specification without output controls, but this differential disappears when adding output measures to the specification. Thus, within Carnegie classifications, public institutions with a smaller state share are not more efficient than institutions with a larger state share.

Of course, comparisons across institutions are always difficult due to the large number of unobserved factors that may influence efficiency. A clearer picture may emerge if we compare the change in an institution's inefficiency between 1991 and 1995 with the change in the institution's state share. Assuming the unobserved factors (e.g., quality) did not vary across the five-year time span, examining changes within institutions reduces the impact of unobservables on the estimated relationship. Such an analysis provides a better and more direct test of the effects of state share reductions on an institution's efficiency. This part of the analysis is performed using a two-step analysis with the institution's excess costs estimated in the first step as the difference between actual costs and the predicted minimum cost from the frontier. The second step involves regressing inefficiency on a vector of variables that includes the state share. Of course, these results are subject to the limitations in the two-step approach discussed earlier. This inefficiency measure is defined as one plus excess cost as a proportion of the minimum cost:

$$\text{Excess cost ratio}_{it} = \frac{(X_{it}\beta + u_{it})}{X_{it}\beta} = 1 + \frac{u_{it}}{X_{it}\beta} \tag{5}$$

where $X_{it}\beta$ represents the predicted minimum cost and u_{it} the excess costs. The predicted minimum cost already takes institution size into account. Thus, the change in the excess cost ratio (excess cost ratio in 1995 minus excess cost ratio in 1991) can be regressed on the change in state share (state share in 1995 minus state share in 1991) without further controls for institution size:

$$\Delta \text{ excess cost ratio}_I = \beta_0 + \beta_1 \, \Delta\text{state share}_i + \varepsilon_i \tag{6}$$

The state share coefficient (β_1) is negative (–4.64) and significant ($t = 3.7$). However, this should not necessarily be interpreted as suggesting that decreases in the state share increased inefficiency. As can be seen in Figure 2, it is possible that institutions with a smaller state share became less inefficient. As discussed earlier, 406 of the 440 institutions had declines in their state share of revenues. Thus, the vast majority of observations are in quadrants II and III of the graph. Of these 406, 273 or 67% became more efficient (quadrant III) between 1991 and 1995. A more accurate interpretation of the regression result is that those states with smaller declines in their state share had greater improvements in ef-

TABLE 5

Stochastic Frontier Estimation Results—Panel Data by Carnegie Classification

	Research		Doctoral		Master's		Bachelor's	
State share	−153.8	−6.835	−0.3350	−1204.3*	−3213.8**	−0.0210	−0.4041	−0.0893
	(0.70)	(1.13)	(0.33)	(1.91)	(8.65)	(0.02)	(0.24)	(0.09)
Other Controls	no	yes	no	yes	no	yes	no	yes
No. of institutions	82		64		231		63	
No. of observations	410		320		1155		315	

NOTES: Two sets of simultaneous equations are estimated for each Carnegie classification with only the inefficiency specification results reported.. The first specification does not control for other factors, while the second specification includes a vector of control variables. The other controls are undergraduate and graduate enrollments, research expenditures, and revenues. Both specifications include an intercept. Each institution is in the sample for five years (1991–1995), so there are five observations per institution. *t*-statistics are in parentheses.

$^{**}p < 0.05$. $^{*}p < 0.10$.

ficiency compared to states with larger state share declines. Indeed, institutions with the largest state share declines tended to become less efficient. Also, it is interesting to note that the same percentage of institutions with increased state shares (23 of the 34 institutions or 67%) became more efficient.

The excess cost ratio is adjusted for institution size, but this does not necessarily account for *changes* in tuition revenues and state appropriations. Changes in tuition revenue and state appropriations are added to the specification in equation (6). Increased tuition revenue is significantly correlated with increased inefficiency, however, changes in state appropriations did not alter the institution's cost efficiency. Given that tuition revenue increased for almost all institutions and that most schools became more efficient over the time period, the results suggest that those institutions with smaller tuition increases had larger efficiency gains. The added variables have virtually no impact on the state share coefficient.

Based on these results, states need to be cautious about simply reducing appropriations and altering universities' revenue structure as a reaction to perceived inefficiency. Institutions with greater declines in state shares had smaller efficiency gains than schools with smaller state share

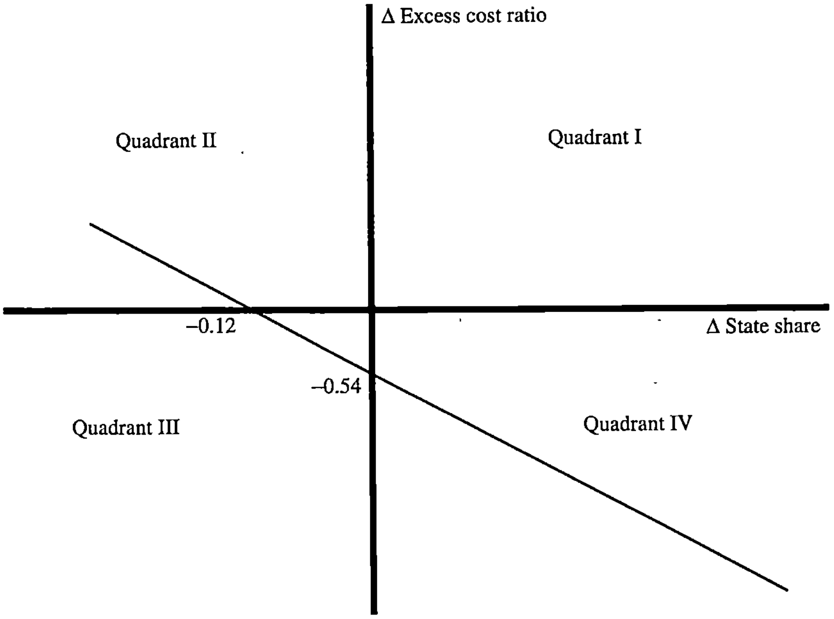


FIG. 2. The Relationship Between the Changes in State Share and the Excess Cost Ratio

declines. Institutions that increased tuition revenues to offset state appropriations reductions tended to become less efficient. Of course, no single state or institution should assume that the results in this article apply to their situation. This article's key policy implication is that legislators need to carefully consider the implications of reducing state appropriations, including how revenues from other sources are likely to change before claiming that appropriation reductions will create more efficient colleges and universities.

As is usually the case, several caveats apply to the results presented in this article. In particular, it should be emphasized that we do not advocate reducing expenditures by "excess" costs. On average, costs may be above the frontier due to factors not taken into account in this study. In addition, as implied throughout this article, such cuts may not lead to the desired results. Also, as discussed above, all of the outputs from higher education cannot be controlled for, and the inefficiency estimate is likely to be measured with some error.

One such exclusion is the lack of quality controls. It is likely that expenditures are positively correlated with institution quality, and such expenditures on quality partly account for "excess" costs. However, this article's conclusions are unlikely to change if quality controls were included. The logic for the cross-sectional results in Tables 3-5 is as follows. Institution quality increases with state appropriations per FTE student (Volkwein, 1986a). To the extent that state appropriations are positively correlated with state share (the correlation is admittedly low, but positive and statistically significant at 0.22 for the 2200 observations), a higher state share should be positively correlated with quality. If excess costs are actually denoting higher quality, the omitted variable bias would lead us to incorrectly find that institutions with a higher state share have greater inefficiency. In other words, if the state share variable was actually capturing the effect of quality, then the estimated relationship between an institution's inefficiency and the state share would be positive, not negative. The argument for the time-series results is more straightforward. As discussed earlier, institutional quality changes relatively slowly, and assuming quality remains constant between 1991 and 1995 (a typical assumption in difference models), the results comparing the change in the excess cost rate with state share changes will be unaffected by quality considerations.

Conclusion

This article had several goals. First, the reduced importance of state appropriations and the increased importance of tuition revenue in public universities was examined. The proportion of revenues from tuition increased during the sample period for the vast majority of universities due to tuition revenue growth combined with either reductions or limited growth in state appropriations. Many states reduced real higher education appropriations in the early 1990s, reducing the share of revenues from state appropriations. This article used a frontier estimation approach to analyze whether the changing revenue structure of institutions affected their cost efficiency. In particular, we considered whether institutions with smaller state shares were more efficient than universities with higher state shares. Using IPEDS data, institutions with smaller state shares were not more efficient than institutions with higher state shares. Last, we examined changes in institutions' state shares and efficiency between 1991 and 1995. Most institutions' state share of revenues fell, but institutions with smaller state share declines increased efficiency more than institutions with larger state share declines. These results have potentially important policy implications given governors and legislators use of appropriations in attempts to influence universities' efficiency.

Notes

[1] Without considering efficiency issues in the sense pursued here, previous studies on state higher education appropriations include Clotfelter (1976), Peterson (1976), Coughlin and Erekson (1986), Hoenack and Pierro (1990), and Strathman (1994).

[2] No attempt is made to rigorously explain why there may be a relationship between state appropriations and efficiency, rather the question is whether state leaders are correct in assuming such a relationship exists. The literature review discusses several articles that imply state appropriations and efficiency might be related.

[3] A number of studies also estimate frontier production functions (see Carlson, 1972; Sengupta & Sfeir, 1986; Cooper & Cohn, 1997).

[4] The cross-sectional methodology used in this article has a long tradition in economics. Typically referred to as the hedonic function or implicit price approach, economists assume that causality can be inferred from cross-sectional results. Those unwilling to accept this assumption can view the cross-sectional results as descriptive.

[5] This argument is similar to the flypaper effect often discussed in the public economics literature. See Feldstein (1978), Craig and Inman (1982), and Tsang and Levin (1983) for examples of how federal grants and aid influence local school district spending.

[6] All dollar values are deflated to 1991 dollars using the Higher Education Price Index (HEPI).

[7] General and educational expenditures include outlays for instruction, research, public service, academic support, student services, institutional support, and maintenance.

[8] Obviously there are a number of additional revenue sources, such as external grants, that may influence tuition charges. Our focus is on the trade-off between state appropriations and tuition revenue, thus we only consider these two revenue sources. Also, external grants comprise a relatively small portion of colleges' total revenues.

[9] Aigner et al. (1977) find the distribution of u_{it} has little impact on the estimated coefficients.

[10] Cohn et al. (1989) consider a sample of research universities for which research output measures are available. They find a 0.7 correlation between research expenditures

and research output. I examine the relationship between research expenditures and the 1995 National Research Council's (NRC) rankings of graduate programs. Webster and Skinner (1996) report the mean score of all programs for 104 institutions with 15 or more programs rated. There exists a 0.76 correlation between research expenditures in 1995 and the mean NRC scores for the 104 graduate programs. Although these results suggest expenditures may be a reasonable proxy for output, De Groot et al. (1991) find that their results are sensitive to whether research expenditures or research output is used in the analysis.

[11] Estimation is performed using FRONTIER Version 4.1. FRONTIER is a computer program that estimates stochastic frontiers using panel data and performs simultaneous equations estimation. The program and all documentation are available on- line at http://www.une.edu.au/econometrics/cepa.htm.

[12] Again the dependent variable is total expenditures. Although expenditures per unit of output would be preferable, colleges and universities have multiple outputs, making the definition of a unit of output unclear. The marginal effects are calculated by taking the derivative of the cost function with respect to undergraduate students, graduate students, or research expenditures. The derivative is evaluated at the mean output levels.

[13] A specification was estimated with state appropriations and tuition revenue included as separate variables with both positively (and significantly) related to inefficiency.

References

Aigner, D. C., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6,* 21–38.

Adams, C. R., Hankins, R. L. & Schroeder, R. G. (1978). *A study of cost analysis in higher education.* Washington, DC: American Council on Education.

Battese, G. E., & Coelli, T. J. (1993). *A stochastic frontier production function incorporating a model for technical inefficiency effects.* Unpublished manuscript. University of New England, Centre for Efficiency and Productivity Analysis, Armidale, New South Wales, Australia.

Battese, G. E., & Coelli, T. J. (1995). A model for technical efficiency effects in a stochastic frontier production function for panel data. *Empirical Economics, 20,* 325-332.

Brinkman, P. T. (1981). Factors affecting instructional costs at major research universities. *Journal of Higher Education, 52,* 265-279.

Brinkman, P. T., & Leslie, L. L. (1986). Economies of scale in higher education: Sixty years of research. *Review of Higher Education, 10,* 1–28.

Brovender, S. (1974). On the economics of a university: Toward the determination of marginal cost of teaching services. *Journal of Political Economy, 82,* 657-664.

Carlson, D. E. (1972). *The production and cost behavior of higher education institutions.* Berkeley, CA: Ford Foundation.

Chronicle of Higher Education. (1999). *The 1999–2000 Almanac Issue.* Washington, DC: Author.

Clotfelter, C. T. (1976). Public spending for higher education: An empirical test of two hypotheses. *Public Finance, 31,* 177-194.

Coelli, T. J. (1996). *A guide to Frontier version 4.1: A computer program for stochastic frontier production and cost function estimation.* Unpublished manuscript. University of New England, Centre for Efficiency and Productivity Analysis, Armidale, New South Wales, Australia.

Coelli, T. J., Rao, D.S., & Battese, G. (1997). *An introduction to efficiency and productivity analysis.* Boston: Kluwer Academic Publishers.

Cohn, E., Rhine, S. L. W., & Santos, M. C. (1989). Institutions of higher education as multi-product firms: Economies of scale and scope. *Review of Economics and Statistics, 71,* 284–290.

Cooper, S. T., & Cohn, E. (1997). Estimation of a frontier production function for the South Carolina educational process. *Economics of Education Review, 16,* 313–327.

Coughlin, C. C., & Erekson, O. H. (1986). Determinants of state aid and voluntary support of higher education. *Economics of Education Review, 5,* 179–190.

Craig, S. G., & Inman, R. P. (1982). Federal aid and public education: An empirical look at the new fiscal federalism. *Review of Economics and Statistics, 64,* 541–551.

De Groot, H., McMahon, W. W., & Volkwein, J. F. (1991). The cost structure of American research universities. *Review of Economics and Statistics, 73,* 424–431.

Dundar, H., & Lewis, D. R. (1995). Departmental productivity in American universities: Economies of scale and scope. *Economics of Education Review, 14,* 119–144.

Feldstein, M. (1978). Effect of a differential add-on grant: Title I and local education spending. *Journal of Human Resources, 13,* 443–458.

Gamkhar, S., & Oates, W. (1996). Asymmetries in the response to increases and decreases in intergovernmental grants: Some empirical findings. *National Tax Journal, 49,* 501–512.

Healy, P., & Schmidt, P. (1997). At public colleges, the tuition debate is a mix of philosophy and practicality. *Chronicle of Higher Education, 43*(38), pp. A17-A18.

Hoenack, S. A., & Pierro, D. J. (1990). An econometric model of a university's income and enrollments. *Journal of Economic Behavior and Organization, 14,* 403–423.

Hoenack, S. A., Weiler, W. C., Goodman, R. D., & Pierro, D. J. (1986). Marginal costs of instruction. *Research in Higher Education, 24,* 335–417.

James, E. (1978). Product mix and cost disaggregation. *Journal of Human Resources, 13,* 157-186.

Koshal, R. K., & Koshal, M. A. (1995). Quality and economies of scale in higher education. *Applied Economics, 27,* 773–778.

McPherson, M. S., & Schapiro, M. O. (1994). The effect of government financing on the behavior of colleges and universities. In M. S. McPherson, M. O. Schapiro, & G. C. Winston (Eds.). *Paying the piper: Productivity, incentives, and financing in U.S. higher education* (pp. 235–250). Ann Arbor: The University of Michigan Press.

Peterson, R. G. (1976). Environmental and political determinants of state higher education appropriations policies. *Journal of Higher Education, 47,* 523–542.

Schmidt, P. (1997). Rancor and confusion greet change in South Carolina budgeting. *Chronicle of Higher Education, 43*(30), pp. A26–A27.

Sengupta, J. K., & Sfeir, R. E. (1986). Production frontier estimates of scale in public schools in California. *Economics of Education Review, 5,* 297–307.

Strathman, J. G. (1994). Migration, benefit spillovers, and state support of higher education. *Urban Studies, 31,* 913–920.

Tierney, M. (1980). An estimate of departmental cost functions. *Higher Education, 9,* 453–468.

Tsang, M., & Levin, H. M. (1983). The impact of intergovernmental grants on educational expenditure. *Review of Educational Research, 53,* 329–367.

Verry, D. W., & Layard, P. R. G. (1975). Cost functions for teaching and research. *Economic Journal, 85,* 55–74.

Volkwein, J. F. (1986a). Campus autonomy and its relationship to measures of university quality. *Journal of Higher Education, 57,* 510–528.

Volkwein, J. F. (1986b). State financial control of public universities and its relationship to campus administrative elaborateness and cost. *Review of Higher Education, 9,* 267–286.

Volkwein, J. F. (1987). State regulation and campus autonomy. In J. C. Smart (Ed.). *Higher education: Handbook of theory and research* (Vol. 3, pp. 120–154). New York: Agathon Press.

Webster, D. S., & Skinner, T. (1996). Rating PhD programs: What the NRC report says . . . and doesn't say. *Change, 28*(3), 22–43.

Book Reviews

Why Universities Matter: A Conversation About Values, Means, and Directions, edited by Tony Coady. St. Leonards, Australia: Allen & Unwin Pty Ltd, 2000. 254+ pp. $24.95

MARC CUTRIGHT, Johns Hopkins University

This recent Australian book is a work that wears controversy most literally on its sleeve, its cover is a collage of a university architecture under dark clouds, and ripped headlines such as "When academics are afraid to speak out," and "How Melbourne Uni muzzled mild dissent."

Why Universities Matter, edited by Tony Coady, director of the Centre for Philosophy and Public Issues at the University of Melbourne, grew out of papers developed for workshops and that Centre. The collection was submitted to the Melbourne University Press (MUP) for publication in the "Ethics and Public Life" series that Coady edits for the press. Although staff reviewers endorsed the book, the management board of the press took the unusual step of rejecting staff advice and the book. Lack of marketability and scant presentation of new research were among the reasons for rejection offered by the board. However, the authors and their champions contended that it was the works' criticisms of managerialism in higher education generally, and of Melbourne University specifically, that were at the heart of the rejection and thus the rejection constituted censorship. Allen & Unwin Press picked up the book and has used the controversy to the benefit of sales.

But more on the book itself. The volume is a collection of twelve essays, three of them commissioned after MUP's rejection. Most authors are respected Australian academicians, although journalist Morag Fraser contributes a concluding chapter on the publication controversy itself.

Coady reflects the tone of most pieces in his introductory chapter, "Universities and the Ideals of Inquiry." The setting here and throughout are the changes that have been wrought in Australian universities through the '80s and '90s. In Australia, as elsewhere, recent decades have been marked by government policies of expanding postsecondary enrollment even while stabilizing or even cutting public financial support. Governments have attempted, by direct or indirect means, to steer institutional enrollments, research, and administrative structures toward consonance with what the governments assert to be national economic interests and global realities.

Coady asserts several effects of a shifting from education to training and the ascendancy of a "cult of management" (p. 16). These include expanding bureaucracy, with attendant "centralist surveillance, and a childish belief in the intrinsic value of collecting social science statistics" (p. 17); a decline in consultation and an increase in authoritarianism in decision making; and an "erosion of the specific rights and privileges of academics," such as tenure, "that were once thought to go with their distinctive vocation and its importance for the community" (p. 21). These and other of Coady's points are consonant with the "worst-case" scenario posited by Slaughter and Leslie in their 1996 *Academic Capitalism.*

Non-Australian readers might do best to start with the chapter by Stuart Macintyre and Simon Marginson, essentially a history of Australian universities with a focus on late decades. Not that this is a for-background-only essay; the authors conclude that the "failure of university administrators to defend the civic university from assault by Commonwealth and state ministers is conspicuous" (p. 68). They hold that the historic purposes of universities in Australia have been to preserve and advance knowledge, prepare students for careers through the fostering of inquiry, and to reflect and comment on public issues. Although the sector has not always accomplished these ends with distinction, "it would be a pity if those who are its present custodians betrayed their responsibilities" (p. 69) to maintain these roles through forfeit to direct commercialization.

Philosopher Raimond Gaita contributes a particularly broad-ranging and strong piece, "Truth and the University," in which he addresses, among other topics, the vocation of teaching and the folly of predicting with specificity what today's students need for "the future" beyond a sound values base and the habit of inquiry. His most pointed commentary, however, is aimed at "Managerial Newspeak and the Corporate University" (p. 46), and the unwavering but undefined allegiance to "excellence" which results in a "pervasive mendacity" that distorts public presentation and suppresses expression of doubt.

Other authors serve up variations and expansions on this criticism. Historian John Molony gives some details on the swelling of administrative offices, in number and in ratio to academic appointments. Vice-chancellors, the heads of campuses, have become "Chief Executive Officers," and have appointed "a phalanx . . . of lesser lights," to which they turn for advice on a painful decision, "after which pronouncements are made, piously but falsely, that a proper process of consultation has been followed" (p. 83). Philosopher Bruce Langtry considers the economic models that drive managerial restructuring of higher education. He notes that other social institutions and contracts, such as marriage, yield to analyses through economic lenses. But just as marriage partners would be unwise to consider their union primarily as a mechanism for exchange of goods and services, so it is a "misuse of economics and management theory to be overly occupied with the economic contribution that universities make to society, or with the economic analyses of their internal functioning" (p. 93). Marginson's sole-author essay holds that governmental and institutional managerialism in research leads to the undesirable homogenization of that research, a focus on quantity over quality, and a favoring of research with immediate economic benefit. Other strong chapters are unnoted here only in concession to brevity; there is the matter of publication controversy to explore further.

To explain its rare contradiction of its professional staff, the board of the Melbourne University Press left few buttons unpressed in its rejection of *Why Universities Matter.* Among its justifications, the work is asserted to be similar to 1998's *On the Brink: Australia's Universities Confronting Their Future,* by Peter Coaldrake and Lawrence Stedman, the sales of which the MUB board characterized as disappointing. Morag Fraser's essay in the present volume holds that the board reached its decision without actually reading *Why Univer-*

sities Matter, but it seems even clearer that they did not read Coaldrake and Stedman, as the comparison seems spurious. Though a well-researched and thoughtful book, *On the Brink* generally supports an expanded role for central management in modern university life.

The board of MUP also said that the articles of *Why Universities Matter* were not derived from "new or recent research." Perhaps; most authors choose to comment from personal perspectives rather than from extensive reference or citation of data. But this is disqualified in toto as scholarship only if one is allegiant to the chestnut that stealing from one source is plagiarism, while stealing from two is research. It is more appropriate to look at these essays as the authors' testimonies about their professional lives and times, to view them as important, primary sources. These are leading academics, and they are disheartened by the directions of their professions, their institutions, and the society they reflect. The considered expression of such ideas was once regarded as central to the university's reason for being.

If there is a disappointment in the book, it comes from what is missing: representation or commentary from the academic field of higher education. With the notable exception of Simon Marginson, a leading scholar of higher education in Australia, the essays were contributed by historians, sociologists, philosophers, physicists, political scientists, and administrators. Although this disciplinary range is a strength of a sort, it does make more apparent the nonparticipation of those who study, write on, and teach about higher education for a living. Sheila Slaughter (1990) noted the absence of American higher education scholars from criticisms of the enterprise, and attributed this circumstance to their focus on institutional functionalism in an era of postsecondary expansion. There have been a few notable, research-based works since then coming from the field of higher education and critical theory, or merely critical, perspectives (e.g., Slaughter & Leslie, 1996; Tierney & Bensimon, 1996), but the pendulum has hardly swung to the opposite extreme. The issues of *Why Universities Matter* are not, of course, unique to Australia, but increasingly international. Perhaps if the field of higher education were to put more research muscle behind the kinds of criticisms offered in the book of present discussion, the boards of presses and other central administrators would pause occasionally in their assertions of the inevitability and desirability of current, questionable directions in university life and purpose.

References

Coaldrake, P., & Stedman, L. (1998). *On the brink: Australia's universities confronting their future.* Queensland, Australia: Queensland University Press.

Slaughter, S. (1990). *The higher learning & high technology.* Albany: State University of New York Press.

Slaughter, S., & Leslie, L. (1996). *Academic capitalism.* Baltimore: Johns Hopkins University Press.

Tierney, W. G., & Bensimon, E. M. (1996). *Promotion and tenure.* Albany: State University of New York Press.

Annual Index

Volume 72, Numbers 1–6 (2001)

1–119	January/February
121–263	March/April
265–385	May/June
387–507	July/August
509–627	September/October
629–758	November/December

Aligning Faculty Rewards with Institutional Mission: Statements, Policies, and Guidelines, Robert M. Diamond, reviewed, 72:1, 112–115

Altbach, Philip G., book review, 72:2, 254–256

Anderson, Melissa S., The Complex Relations Between the Academy and Industry: Views from the Literature, 72:2, 226–246

Annual Index, 72:6, 754–758

Antony, James Soto, book review, 72:3, 372–374

Assessment in Higher Education: Issues of Access, Quality, Student Development, and Public Policy, Samuel J. Messick (Ed.), reviewed, 72:1, 117–119

Baird, Leonard, The Social Role of Higher Education: Introduction to a Special *JHE* Issue, 72:2, 121–123

Barak, Rachel, *see* Hativa, Nira

Barr, Carol A., *see* Covell, Dan

Bayer, Alan E., book review, 72:1, 115–117

Berger, Joseph B., book review, 72:1, 112–115

Bianco-Mathis, Virginia, and Neal Chalofsky (Eds.), *The Full-Time Faculty Handbook,* reviewed, 72:1, 110–112

Bieber, Jeffery P., book review, 72:1, 110–112

Bjorklund, Stefani A., *see* Terenzini, Patrick T.

Bonk, Curtis Jay, and Kira S. King (Eds.), *Electronic Collaborators: Learner-centered Technologies for Literacy, Apprenticeship, and Discourse,* reviewed, 72:3, 374–376

Braxton, John M. (Ed.), *Perspectives on Scholarly Misconduct in the Sciences,* reviewed, 72:1, 108–110

Bringle, Robert G., Richard Games, and Reverend Edward A. Malloy (Eds.), *Colleges and Universities as Citizens,* reviewed, 72:2, 247–249

Brown II, M. Christopher, Collegiate Desegregation and the Public Black College: A New Policy Mandate, 72:1, 46–62

Bryan, Glenn A., *see* Ghosh, Amit K.

Building the Responsive Campus: Creating High Performance Colleges and Universities, William G. Tierney, reviewed, 72:2, 249–253

Butterfield, Kenneth D., *see* McCabe, Donald L.

Cabrera, Alberto F., *see* Terenzini, Patrick T.

Career Strategies for Women in Academe: Arming Athena, Lynn H. Collins, Joan C. Chrisler, and Kathryn Quina (Eds.), reviewed, 72:1, 101–103

Checkoway, Barry, Renewing the Civic Mission of the American Research University, 72:2, 125–147

Chism, Nancy Van Note, *The Peer Review of Teaching: A Sourcebook,* reviewed, 72:5, 612–619

Coady, Tony (Ed.), *Why Universities Matter: A Conversation About Values, Means, and Directions,* reviewed, 72:6, 751–753

Colbeck, Carol L., *see* Terenzini, Patrick T.

Colleges and Universities as Citizens, Robert G. Bringle, Richard Games, and Reverend Edward A. Malloy (Eds.), reviewed, 72:2, 247–249

Collegiate Desegregation and the Public Black College: A New Policy Mandate, M. Christopher Brown II, 72:1, 46–62

Collins, Lynn H., Joan C. Chrisler, and Kathryn Quina (Eds.), *Career Strategies for Women in Academe,* reviewed, 72:1, 101–103

Complex Relations Between the Academy and Industry, The: Views from the Literature, Melissa S. Anderson, 72:2, 226–246

Condemning Students to Debt: College Loans and Public Policy, Richard Fossey and Mark Bateman (Eds.), reviewed, 72:4, 506–507

Cost Efficiency in Public Higher Education Institutions, John Robst, 72:6, 730–750
Covell, Dan, and Carol A. Barr, The Ties That Bind: Presidential Involvement with the Development of NCAA Division I Initial Eligibility Legislation, 72:4, 414–452
Creamer, Don G., book review, 72:5, 620–625
Currie, Jan, and Janice Newsom (Eds.), *Universities and Globalization: Critical Perspectives,* reviewed, 72:2, 254–256
Cutright, Marc, book review, 72:6, 751–753

De Valero, Yaritza Ferrer, Departmental Factors Affecting Time- to-Degree and Completion Rates of Doctoral Students at One Land-Grant Research Institution, 72:3, 341–367
Denison, D. Brian, *see* Donald, Janet G.
Departmental Factors Affecting Time-to-Degree and Completion Rates of Doctoral Students at One Land-Grant Research Institution, Yaritza Ferrer de Valero, 72:3, 341–367
Diamond, Robert M., *Aligning Faculty Rewards with Institutional Mission: Statements, Policies, and Guidelines,* reviewed, 72:1, 112–115
Dickeson, Robert C., *Prioritizing Academic Programs and Services: Reallocating Resources to Achieve Strategic Balance,* reviewed, 72:5, 622–625
Dishonesty in Academic Environments: The Influence of Peer Reporting Requirements, Donald L. McCabe, Linda K. Trevino, and Kenneth D. Butterfield, 72:1, 29–45
Donald, Janet G., and D. Brian Denison, Quality Assessment of University Students: Student Perceptions of Quality Criteria, 72:4, 478–502
Dynamic Tensions of Service Learning in Higher Education, The: A Philosophical Perspective, Adrianna Kezar, and Robert A. Rhoads, 72:2, 148–171

Edison, Marcia I., *see* Whitt, Elizabeth J.
Einarson, Marne K., *see* Peterson, Marvin W.
Electronic Collaborators: Learner-centered Technologies for Literacy, Apprenticeship, and Discourse, Curtis Jay Bonk and Kira S. King (Eds.), reviewed, 72:3, 374–376
Enhancing Student Learning: Setting the Campus Context, Frances Stage, Lemuel W. Watson, and Melvin Terrell (Eds.), reviewed, 72:5, 620–622
Ethington, Corinna A., *see* Feldman, Kenneth A.
Exemplary University Teachers: Knowledge and Beliefs Regarding Effective Teaching Dimensions and Strategies, Nira Hativa, Rachel Barak, and Etty Simhi, 72:6, 699–729
Eyler, Janet, and Dwight E. Giles, Jr., *Where's the Learning in Service-Learning?* reviewed, 72:2, 256–258

Faculty-Student Sexual Involvement: Issues and Interventions, Virginia L. Stamler and Gerald L. Stone, reviewed, 72:1, 106–108.
Feldman, Kenneth A., Corinna A. Ethington, and John C. Smart, A Further Investigation of Major Field and Person-Environment Fit: Sociological Versus Psychological Interpretations of Holland's Theory, 72:6, 670–698
Financing a College Education: How It Works, How It's Changing, Jacqueline E. King (Ed.), reviewed, 72:3, 370–372
Flowers, Lamont, Steven J. Osterlind, Ernest T. Pascarella, and Christopher T. Pierson, How Much Do Students Learn in College?: Cross-Sectional Estimates Using the College BASE, 72:5, 565–583
Forest, James J. F. (Ed.), *University Teaching: International Perspectives,* reviewed, 72:4, 503–506
Fossey, Richard, and Mark Bateman (Eds.), *Condemning Students to Debt: College Loans and Public Policy,* reviewed, 72:4, 506–507
Freedom's Web: Student Activism in an Age of Cultural Diversity, Robert A. Rhoads, reviewed, 72:3, 376–379
Full-Time Faculty Handbook, The, Virginia Bianco-Mathis and Neal Chalofsky (Eds.), reviewed, 72:1, 110–112
Funding Science in America: Congress, Universities, and the Politics of the Academic Pork Barrel, James D. Savage, reviewed, 72:2, 258–260
Further Investigation of Major Field and Person-Environment Fit, A: Sociological Versus Psychological Interpretations of Holland's Theory, Kenneth A. Feldman, Corinna A. Ethington, and John C. Smart, 72:6, 670–698

Geiger, Roger, book review, 72:2, 258–260
Ghosh, Amit K., Thomas W. Whipple, and Glenn A. Bryan, Student Trust and Its Antecedents in Higher Education, 72:3, 322–340
Glazer-Raymo, Judith, *Shattering the Myths: Women in Academe,* reviewed, 72:5, 625–627
Globalization of Higher Education, The, Peter Scott (Ed.), reviewed, 72:2, 254–256
Going to College: How Social, Economic, and Educational Factors Influence the Decisions Students Make, Don Hossler, Jack Schmit, and Nick Vesper, reviewed, 72:3, 383–385
Grubb, W. Norton & Associates (Eds.), *Honored but Invisible: An Inside Look at Teaching in Community Colleges,* reviewed, 72:3, 379–383

Hamrick, Florence A., book review, 72:5, 625–627
Hativa, Nira, Rachel Barak, and Etty Simhi, Exemplary University Teachers: Knowledge

and Beliefs Regarding Effective Teaching Dimensions and Strategies, 72:6, 699–729
Heller, Donald E., book review, 72:3, 370–372
Hermanowicz, Joseph C., *The Stars Are Not Enough: Scientists–Their Passions and Professions,* reviewed, 72:3, 372–374
Hernon, Peter, book review, 72:1, 108–110
Honored but Invisible: An Inside Look at Teaching in Community Colleges, W. Norton Grubb & Associates (Eds.), reviewed, 72:3, 379–383
Hossler, Don, Jack Schmit, and Nick Vesper, *Going to College: How Social, Economic, and Educational Factors Influence the Decisions Students Make,* reviewed, 72:3, 383–385
How Much Do Students Learn in College?: Cross-Sectional Estimates Using the College BASE, Lamont Flowers, Steven J. Osterlind, Ernest T. Pascarella, and Christopher T. Pierson, 72:5, 565–583
Hu, Shouping, *see* Kuh, George D.
Hu, Shouping, and Edward P. St. John, Student Persistence in a Public Higher Education System: Understanding Racial and Ethnic Differences, 72:3, 265–286
Hutcheson, Philo A., *A Professional Professoriate: Unionization, Bureaucratization, and the AAUP,* reviewed, 72:3, 368–369
Hyle, Adrienne E., *see* Mills, Michael R.

Influences on Students' Openness to Diversity and Challenge in the Second and Third Years of College, Elizabeth J. Whitt, Marcia I. Edison, Ernest T. Pascarella, Patrick T. Terenzini, and Amaury Nora, 72:2, 172–204
Institutional Effectiveness of Women-Only Colleges: Cultivating Students' Desire to Influence Social Conditions, Mikyong Minsun Kim, 72:3, 287–321

Jacoby, Barbara, book review, 72:2, 247–249
Jones, Elizabeth A., book review, 72:1, 117–119
Jones, Susan R., book review, 72:2, 256–258
Jun, Alexander, *see* Tierney, William G.

Keig, Larry, The Role of Faculty in Assessing Colleagues' Teaching, review essay, 72:5, 612–619
Kezar, Adrianna, and Robert A. Rhoads, The Dynamic Tensions of Service Learning in Higher Education: A Philosophical Perspective, 72:2, 148–171
Kim, Mikyong Minsun, Institutional Effectiveness of Women-Only Colleges: Cultivating Students' Desire to Influence Social Conditions, 72:3, 287–321
King, Jacqueline E. (Ed.), *Financing a College Education: How It Works, How It's Changing,* reviewed, 72:3, 370–372
Kovalchik Ann, book review, 72:3, 374–376
Kuh, George D., and Shouping Hu, Learning Productivity at Research Universities, 72:1, 1–28

Learning Productivity at Research Universities, George D. Kuh and Shouping Hu, 72:1, 1–28

Magolda, Peter M., book review, 72:5, 620–622
Manning, Kathleen, book review, 72:3, 376–379
McCabe, Donald L., Linda K. Trevino, and Kenneth D. Butterfield, Dishonesty in Academic Environments: The Influence of Peer Reporting Requirements, 72:1, 29–45
Melko, Matthew, *A Professor's Work,* reviewed, 72:1, 115–117
Messick, Samuel J. (Ed.), *Assessment in Higher Education: Issues of Access, Quality, Student Development, and Public Policy,* reviewed, 72:1, 117–119
Mills, Michael R., and Adrienne E. Hyle, No Rookies on Rookies: Compliance and Opportunism in Policy Implementation, 72:4, 453–477
Modeling Supply and Demand for Arts and Sciences Faculty: What Ten Years of Data Tell Us About the Labor Market Projections of Bowen and Sosa, Douglas T. Shapiro, 72:5, 532–564
Models of Institutional Resource Allocation: Mission, Market, and Gender, Cindy S. Volk, Sheila Slaughter, and Scott L. Thomas, 72:4, 387–413
Moore, Kathryn M., book review, 72:2, 249–253

No Neutral Ground: Standing by the Values We Prize in Higher Education, Robert B. Young, reviewed, 72:2, 261–263
No Rookies on Rookies: Compliance and Opportunism in Policy Implementation, Michael R. Mills and Adrienne E. Hyle, 72:4, 453–477
Nora, Amaury, *see* Whitt, Elizabeth J.

Olivas, Michael A., book review, 72:3, 368–369
Osterlind, Steven J., *see* Flowers, Lamont

Parente, John M., *see* Terenzini, Patrick T.
Pascarella, Ernest T., *see* Whitt, Elizabeth J.
——— *see* Flowers, Lamont
Paulsen, Michael B., book review, 72:3, 383–385
Peer Review of Teaching, The: A Sourcebook, Nancy Van Note Chism, reviewed, 72:5, 612–619
Perna, Laura W., The Relationship Between Family Responsibilities and Employment

Status Among College and University Faculty, 72:5, 584–611

Perspectives on Scholarly Misconduct in the Sciences, John M. Braxton (Ed.), reviewed, 72:1, 108–110

Peterson, Marvin W., and Marne K. Einarson, What Are Colleges Doing About Student Assessment: Does It Make a Difference? 72:6, 629–669

Pierson, Christopher T., *see* Flowers, Lamont

Prioritizing Academic Programs and Services: Reallocating Resources to Achieve Strategic Balance, Robert C. Dickeson, reviewed, 72:5, 622–625

Professional Professoriate, A: Unionization, Bureaucratization, and the AAUP, Philo A. Hutcheson, reviewed, 72:3, 368–369

Professor's Work, A, Matthew Melko, reviewed, 72:1, 115–117

Quality Assessment of University Students: Student Perceptions of Quality Criteria, Janet G. Donald and D. Brian Denison, 72:4, 478–502

Racial and Ethnic Diversity in the Classroom: Does It Promote Student Learning? Patrick T. Terenzini, Alberto F. Cabrera, Carol L. Colbeck, Stefani A. Bjorklund, and John M. Parente, 72:5, 509–531

Relationship Between Family Responsibilities and Employment Status Among College and University Faculty, The, Laura W. Perna, 72:5, 584–611

Renewing the Civic Mission of the American Research University, Barry Checkoway, 72:2, 125–147

Rhoads, Robert A., *Freedom's Web: Student Activism in an Age of Cultural Diversity,* reviewed, 72:3, 376–379

——— *see* Kezar, Adrianna

Robst, John, Cost Efficiency in Public Higher Education Institutions, 72:6,730–750

Role of Faculty in Assessing Colleagues' Teaching, The, Larry Keig, review essay, 72:5, 612–619

Rupert, Patricia A., book review, 72:1, 106–108

St. John, Edward P., *see* Hu, Shouping

Savage, James D., *Funding Science in America: Congress, Universities, and the Politics of the Academic Pork Barrel,* reviewed, 72:2, 258–260

Sax, Linda J., book review, 72:1, 101–103

Scott, Peter (Ed.), *The Globalization of Higher Education,* reviewed, 72:2, 254–256

Segregated by Subject: Racial Differences in the Factors Influencing Academic Major Between European Americans, Asian Americans, and African, Hispanic, and Native Americans, Jacqueline C. Simpson, 72:1, 63–100

Shapiro, Douglas T., Modeling Supply and Demand for Arts and Sciences Faculty: What Ten Years of Data Tell Us About the Labor Market Projections of Bowen and Sosa, 72:5, 532–564

Shattering the Myths: Women in Academe, Judith Glazer-Raymo, reviewed, 72:5, 625–627

Simhi, Etty, *see* Hativa, Nira

Simpson, Jacqueline C., Segregated by Subject: Racial Differences in the Factors Influencing Academic Major Between European Americans, Asian Americans, and African, Hispanic, and Native Americans, 72:1, 63–100

Slaughter, Sheila, *see* Volk, Cindy S.

Smart, John C., *see* Feldman, Kenneth A.

Social Role of Higher Education, The: Introduction to a Special *JHE* issue, Leonard L. Baird, 72:2, 121–123

Stage, Frances, Lemuel W. Watson, and Melvin Terrell (Eds.), *Enhancing Student Learning: Setting the Campus Context,* reviewed, 72:5, 620–622

Stamler, Virginia L., and Gerald L. Stone, *Faculty-Student Sexual Involvement: Issues and Interventions,* reviewed, 72:1, 106–108

Stars Are Not Enough, The: Scientists–Their Passions and Professions, Joseph C. Hermanowicz, reviewed, 72:3, 372–374

Student Persistence in a Public Higher Education System: Understanding Racial and Ethnic Differences, Shouping Hu and Edward P. St. John, 72:3, 265–285

Student Trust and Its Antecedents in Higher Education, Amit K. Ghosh, Thomas W. Whipple, and Glenn A. Bryan, 72:3, 322–340

Sundberg, David C., book review, 72:2, 261–263

Sweetland, Scott R., book review, 72:4, 506–507

Terenzini, Patrick T., *see* Whitt, Elizabeth J.

Terenzini, Patrick T., Alberto F. Cabrera, Carol L. Colbeck, Stefani A. Bjorklund, and John M. Parente, Racial and Ethnic Diversity in the Classroom: Does It Promote Student Learning? 72:5, 509–531

Thomas, Scott L., *see* Volk, Cindy S.

Tierney, William G., *Building the Responsive Campus: Creating High Performance Colleges and Universities,* reviewed, 72:2, 249–253

Tierney, William G., and Alexander Jun, A University Helps Prepare Low-Income Youths for College: Tracking School Success, 72:2, 205–225

Ties That Bind, The: Presidential Involvement with the Development of NCAA Division I

Initial Eligibility Legislation, Dan Covell and Carol A. Barr, 72:4, 414–452

Trevino, Linda K., *see* McCabe, Donald L.

Tullberg, Rita McWilliams, *Women at Cambridge,* reviewed, 72:1, 103–105

Twombly, Susan, book review, 72:3, 379–383

Universities and Globalization: Critical Perspectives, Jan Currie and Janice Newsom (Eds.), reviewed, 72:2, 254–256

University Helps Prepare Low-Income Youths for College, A: Tracking School Success, William G. Tierney and Alexander Jun, 72:2, 205–225

University Teaching: International Perspectives, James J. F. Forest (Ed.), reviewed, 72:4, 503–506

Volk, Cindy S., Sheila Slaughter, and Scott L. Thomas, Models of Institutional Resource Allocation: Mission, Market, and Gender, 72:4, 387–413

Walton, Andrea, book review, 72:1, 103–105

What Are Colleges Doing About Student Assessment: Does It Make a Difference? Marvin W. Peterson and Marne K. Einarson, 72:6, 629–669

Where's the Learning in Service-Learning? Janet Eyler and Dwight E. Giles, Jr., reviewed, 72:2, 256–258

Whipple, Thomas W., *see* Ghosh, Amit K.

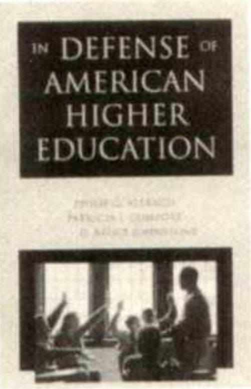
Whitt, Elizabeth J., Marcia I. Edison, Ernest T. Pascarella, Patrick T. Terenzini, and Amaury Nora, Influences on Students' Openness to Diversity and Challenge in the Second and Third Years of College, 72:2, 172–204

Why Universities Matter: A Conversation About Values, Means, and Directions, Tony Coady (Ed.), reviewed, 72:6, 751–753

Wimsatt, Leslie Adams, book review, 72:4, 503–506

Women at Cambridge, Rita McWilliams Tullberg, reviewed, 72:1, 103–105

Young, Robert B., *No Neutral Ground: Standing by the Values We Prize in Higher Education,* reviewed, 72:2, 261–263



In Defense of American Higher Education

edited by Philip G. Altbach, Patricia J. Gumport, and D. Bruce Johnstone

"This is a splendid book that will give readers a panoramic overview of university affairs, among the most important aspects of the American political economy, and its social system and the visions for it."—Martin Meyerson, University Professor Emeritus and President Emeritus, University of Pennsylvania

$19.95 paperback

Higher Education Law

The Faculty

Steven G. Poskanzer

"Amidst all of the books and publications relating to the legal issues for colleges and universities, it is amazing that this is the first book to provide deans, department chairs, and general readers with a clear, concise, thoughtful and well-written overview of the legal principles that govern what colleges and universities can and cannot do."—Richard Zansitis, General Counsel, Rice University

$19.95 paperback

When Women Ask the Questions

Creating Women's Studies in America

Marilyn Jacoby Boxer

foreword by Catharine R. Stimpson

"This book is enormously valuable as a history of the first twenty-five years of women's studies within the larger context of higher education in the United States."—Mari Jo Buhle, *American Historical Review*

$18.95 paperback

The Political University

Policy, Politics, and Presidential Leadership in the American Research University

Robert M. Rosenzweig

with a new afterword by the author

"Rosenzweig combines unabashed admiration for American research universities, honesty about their real-world shortcomings, keen appreciation for academic quality, and apprehension bordering on congenital pessimism about looming threats to their continued effectiveness."—*Annals of the American Academy of Political and Social Science*

$16.95 paperback

Play-by-Play

Radio, Television, and Big-Time College Sport

Ronald A. Smith

"Many authors have written celebrations—or diatribes—about the commercialism of college sports. Smith is more interesting and effective because he evades the polemics and settles for reconstructing and interpreting a fascinating tale."—John R. Thelin, University of Kentucky

$45.00 hardcover

The Johns Hopkins University Press • 1-800-537-5487 • www.jhupbooks.com

THE GOLD AND THE BLUE

A Personal Memoir of the University of California, 1949–1967, Volume 1, Academic Triumphs

Clark Kerr

"This is a major piece of work going far beyond California in interest, and you cannot doubt that I am an ardent supporter. Truly, on this occasion we can say that we will never see his like again."

—John Kenneth Galbraith

One of the most influential figures in higher education, Kerr was a leading visionary, architect, leader, and fighter for the University of California. Chancellor from 1952 to 1958 and president from 1958 to 1967, he saw the university through its golden years—a time of both great advancement and great conflict. This absorbing memoir is an intriguing insider's account of how the University of California rose to the peak of scientific and scholarly stature and how, under Kerr's unique leadership, the university evolved into the institution it is today.

$35.00 at bookstores or order (800) 822-6657 • www.ucpress.edu

UNIVERSITY OF CALIFORNIA PRESS